JCTC Journal of Chemical Theory and Computation

Systematic Approach for Computing Zero-Point Energy, Quantum Partition Function, and Tunneling Effect Based on Kleinert's Variational Perturbation Theory

Kin-Yiu Wong* and Jiali Gao

Department of Chemistry and Digital Technology Center, University of Minnesota, Minneapolis, Minnesota 55455

Received March 29, 2008

Abstract: In this paper, we describe an automated integration-free path-integral (AIF-PI) method [Wong, K.-Y.; Gao, J. *J. Chem. Phys.* **2007**, *127*, 211103], based on Kleinert's variational perturbation (KP) theory, to treat internuclear quantum-statistical effects in molecular systems. We have developed an analytical method to obtain the centroid potential as a function of the variational parameter in the KP theory, which avoids numerical difficulties in path-integral Monte Carlo or molecular dynamics simulations, especially at the limit of zero-temperature. Consequently, the variational calculations using the KP theory can be efficiently carried out beyond the first order, i.e., the Giachetti-Tognetti-Feynman-Kleinert variational approach, for realistic chemical applications. By making use of the approximation of independent instantaneous normal modes (INM), the AIF-PI method can be applied to many-body systems, and it was shown previously that the AIF-PI method is accurate for computing the quantum effects including a water molecule and the collinear H_3 reaction. In this work, the accuracy and properties of the KP theory are further investigated by using the first three-order perturbations on an asymmetric double-well potential, the bond vibrations of H_2, HF, and HCl represented by the Morse potential, and a proton-transfer barrier modeled by the Eckart potential. The zero-point energy, quantum partition function, and tunneling factor for these systems have been determined and are found to be in excellent agreement with the exact quantum results. Using our new analytical results at the zero-temperature limit, we show that the minimum value of the computed centroid potential in the KP theory is in excellent agreement with the ground-state energy (zero-point energy), and the position of the centroid potential minimum is the expectation value of particle position in wave mechanics. The fast convergent property of the KP theory is further examined in comparison with results from the traditional Rayleigh−Ritz variational approach and Rayleigh−Schrödinger perturbation theory in wave mechanics. The present method can be used for thermodynamic and quantum dynamic calculations, including systematically determining the exact value of zero-point energy and studying kinetic isotope effects for chemical reactions in solution and in enzymes.

1. Introduction

Kleinert's variational perturbation (KP) theory[1−7] for the centroid density[8−15] of Feynman path integrals[1,8,16−26] provides a complete theoretical foundation for developing nonstochastic methods[27] to systematically incorporate internuclear quantum-statistical effects[28] in condensed phase systems. Computational methods based on the KP theory complement conventional Fourier or discretized path-integral Monte-Carlo[29−39] (PIMC) and molecular dynamics[12,13,40−43] (PIMD) simulations which have been widely used in condensed phases.[44−58] Recently, we reported a computational method based on the KP theory for chemical applications.[59] In this approach which is called an automated integration-free path-integral (AIF-PI) method,[59] the path integrals in the perturbation expansion have been analytically

* Corresponding author e-mail: kiniu@umn.edu; permanent e-mail: kiniu@alumni.cuhk.net.

10.1021/ct800109s CCC: $40.75 © 2008 American Chemical Society
Published on Web 09/09/2008

integrated, resulting in expressions that are free of path integrations and can be efficiently used to study quantum-statistical effects. For many-body systems, we make use of the independent instantaneous normal mode (INM) approximation such that the internuclear potential energy function along each instantaneous normal mode coordinate is expanded in terms of one-dimensional polynomial functions.

To this end, we derived analytical expressions for the centroid effective potential up to the third order of the Kleinert perturbation (KP3).[59] The most attractive feature of the KP theory is that the perturbation series converges uniformly and exponentially.[1,60,61] We have shown that the second-order KP theory (KP2) implemented in the AIF-PI method with the INM approximation is accurate for a number of test cases, including the quantum partition function of a water molecule (3 degrees of freedom) and the rate of the collinear H_3 reaction (2 degrees of freedom), in comparison with accurate quantum results.[59] Moreover, owing to the integration-free feature, our AIF-PI method is computationally efficient such that the potential energy can be evaluated using *ab initio*[62−64] or density-functional theory[65] (DFT) to perform the so-called *ab initio* path-integral calculations.[44−50] Consequently, we used the hybrid functional B3LYP[66] to construct the potential energy function to compute kinetic isotope effects (KIE) on a series of proton transfer reactions in water with the AIF-PI method. The computed KIE results at the KP2 level are in excellent agreement with experiment.[59]

A closely related theoretical approach is the variational method independently introduced by Giachetti and Tognetti[67] and by Feynman and Kleinert[68] (hereafter labeled as GTFK), which formally corresponds to the first-order approximation in the KP theory, i.e., KP1.[1,67−69] The GTFK method has been applied to a variety of systems,[1,70−79] including quantum dynamic processes in condensed phases (e.g., water and helium).[76−79] Although the original GTFK approach is among the most accurate approximate methods for estimating the path-integral centroid potential in many applications,[27] significant errors can exist in situations in which quantum effects are dominant, especially at low temperatures.[27] Our initial report[59] as well as studies by Kleinert et al. on model systems[1−7] showed that higher order perturbations of KP theory can significantly and systematically improve computational accuracy over the KP1 results.

In this article, we use the AIF-PI method to further examine the computational accuracy and properties of the KP theory, making use of a number of test cases that have been well-characterized analytically and computationally. These include an asymmetric double-well potential,[27] the Morse potential,[80] and the Eckart potential.[81]

For the double-well and Morse potentials, we focus on the quantum partition function as a function of temperature as well as the free energy at the zero-temperature limit ($T = 0$ K), where the minimization of the centroid potential yields two important physical quantities:[1,14,15,68] the exact ground-state energy, i.e., zero-point energy (ZPE), and the expectation value of the nuclear position in the ground state. Hence, the newly derived analytical results at the limit of zero-temperature (Supporting Information) provide a convenient way to systematically compute the exact values of these two important physical quantities without solving the vibrational Schrödinger equation. At the zero-temperature limit, we demonstrate that the fast convergent property of KP theory becomes transparent by comparing the ground-state energy of the Morse potential with that determined by the traditional Rayleigh−Ritz variational method[82−85] and Rayleigh−Schrödinger perturbation theory.[63,86,87] For the Eckart potential, we focus on the tunneling effect[88,89] corresponding to a proton transfer at a wide range of temperatures. Comparison of the AIF-PI method with other approximate methods, PIMC or PIMD simulations, and accurate quantum results are also given. In addition, we discuss the selection of the optimal variational parameter Ω, and the temperature-dependence of zero-Ω limit which corresponds to the free-particle reference frame used in the Feynman-Hibbs variational approach.[8]

2. Kleinert's Variational Perturbation Theory

In this section, we briefly review Kleinert's variational perturbation (KP) theory[1−7] and its relationship to the original Giachetti-Tognetti and Feynman-Kleinert (GTFK) variational approach.[1,67−69] The path-integral (PI) representation of the canonical quantum mechanical (QM) partition function Q_{QM} for a one-particle one-dimensional system can be written in terms of the centroid effective potential W as a classical configuration integral[1,8,13−15]

$$Q_{QM} = \sqrt{\frac{Mk_BT}{2\pi\hbar^2}}\int_{-\infty}^{\infty} e^{-\beta W(x_0)}dx_0 \tag{1}$$

where M is the mass of the particle, k_B is Boltzmann's constant, T is temperature, $\hbar$ is Planck's constant divided by 2π, $\beta = 1/k_BT$, and x_0 is a point in configurational space. Given the centroid potential $W(x_0)$, thermodynamic and quantum dynamic quantities can be accurately determined,[1,8,13−15,44−50] including molecular spectroscopy of quantum fluids[76−79] and the rate constant of chemical and enzymatic reactions.[9−13,51−58,69] The mass-dependent nature of $W(x_0)$ is also of particular interest because isotope effects can be obtained, and it has been applied to enzymatic reactions[52,53,56−58] and carbon nanotubes.[90]

The centroid potential $W(x_0)$ in eq 1 is defined as follows[1,8,13−15]

$$W(x_0) = -k_BT\ln\left[\sqrt{\frac{2\pi\hbar^2}{Mk_BT}}\oint \mathscr{D}[x(\tau)]\delta(\bar{x}-x_0)\exp\{-\mathscr{A}[x(\tau)]/\hbar\}\right] \tag{2}$$

where τ is imaginary time, $x(\tau)$ is a function describing a path in space-time, $\oint\mathscr{D}[x(\tau)]\ \delta(\bar{x}-x_0)$ denotes a summation over *all* possible closed paths in which $\bar{x}$ is equal to x_0 (i.e., a functional integration), and $\bar{x}$ is the time-average position, called 'centroid'

$$\bar{x} \equiv \frac{1}{\beta\hbar}\int_0^{\beta\hbar} x(\tau)d\tau \tag{3}$$

In eq 2, $\mathscr{A}$ is the quantum-statistical action

$$\mathscr{A}[x(\tau)] = \int_0^{\beta\hbar} d\tau\left\{\frac{M}{2}\dot{x}(\tau)^2 + V[x(\tau)]\right\} \tag{4}$$

where $V(x)$ is the potential energy function of the system. Generalization of eqs 1and 2 to a multidimensional system is straightforward.[1,8]

A number of computational approaches have been developed to approximately estimate the centroid potential. Feynman and Hibbs described a first-order cumulant expansion by introducing a Gaussian smearing function in a free-particle reference frame to yield an upper bound on the centroid potential.[8] This was subsequently modified by Doll and Myers (DM) by using a Gaussian width associated with the angular frequency at the minimum of the original potential.[91] The GTFK approach is another variational method that adopts a harmonic reference state by variationally optimizing the angular frequency.[67−69] Mielke and Truhlar employed a free-particle reference state and approximated the sum over paths by a minimal set of paths constrained for a harmonic oscillator. The action integral (eq 4) is then obtained by using the three-point trapezoidal rule for the potential to yield the displaced-point path-integral (DPPI) centroid potential.[27]

In Kleinert's variational perturbation (KP) theory,[1−7] one first constructs a harmonic reference state characterized by a trial angular frequency Ω at a given centroid position x_0 (and temperature T), and then systematically builds up anharmonic corrections to the centroid potential of this reference system. Given the reference, or trial harmonic action

$$\mathscr{A}_\Omega^{x_0} = \int_0^{\beta\hbar} d\tau \left\{ \frac{M}{2}\dot{x}(\tau)^2 + \frac{1}{2}M\Omega^2[x(\tau) - x_0]^2 \right\} \qquad (5)$$

the centroid potential $W(x_0)$ in eq 2 can be expressed as a path integral of the harmonic action which is perturbed by the anharmonicity of the original potential

$$e^{-\beta W(x_0)} = \sqrt{\frac{2\pi\hbar^2}{Mk_BT}} \oint \mathscr{D}[x(\tau)]\delta(\bar{x} - x_0)e^{-\mathscr{A}_\Omega^{x_0}/\hbar}e^{-(\mathscr{A}-\mathscr{A}_\Omega^{x_0})/\hbar}$$
$$= Q_\Omega^{x_0}\langle e^{-(\mathscr{A}-\mathscr{A}_\Omega^{x_0})/\hbar}\rangle_\Omega^{x_0} \qquad (6)$$

where $Q_\Omega^{x_0}$ is the local harmonic partition function given as follows

$$Q_\Omega^{x_0} = \sqrt{\frac{2\pi\hbar^2}{Mk_BT}} \oint \mathscr{D}[x(\tau)]\delta(\bar{x} - x_0)e^{-\mathscr{A}_\Omega^{x_0}/\hbar} = \frac{\beta\hbar\Omega/2}{\sinh(\beta\hbar\Omega/2)} \qquad (7)$$

and $\langle\cdots\rangle_\Omega^{x_0}$ is the expectation value over all closed paths of the harmonic action in eq 5 (i.e., a functional average)

$$\langle e^{-F[x(\tau)]/\hbar}\rangle_\Omega^{x_0} = \frac{1}{Q_\Omega^{x_0}}\sqrt{\frac{2\pi\hbar^2}{Mk_BT}} \oint \mathscr{D}[x(\tau)] \times$$
$$\delta(\bar{x} - x_0)e^{-F[x(\tau)]/\hbar}e^{-\mathscr{A}_\Omega^{x_0}/\hbar} \qquad (8)$$

In eq 8, $F[x(\tau)]$ denotes an arbitrary functional. It is of interest to note that eq 6 is the starting point of Zwanzig's statistical-mechanical perturbation theory,[92] which has been extensively used in free energy calculations through Monte Carlo and molecular dynamics simulations.[93,94]

If we expand the exponential functional in eq 6 and sum up the prefactors into an exponential series of cumulants,[95] the nth-order approximation, $W_n^\Omega(x_0)$, to the centroid potential $W(x_0)$ can be written as follows[1,2]

$$e^{-\beta W_n^\Omega(x_0)} = Q_\Omega^{x_0} \exp\Big\{-\frac{1}{\hbar}\int_0^{\beta\hbar} d\tau\langle\mathscr{A}_{\rm int}^{x_0}\rangle_{\Omega,c}^{x_0} +$$
$$\frac{1}{2!\hbar^2}\int_0^{\beta\hbar} d\tau_1\int_0^{\beta\hbar} d\tau_2\langle\mathscr{A}_{\rm int}^{x_0}[x(\tau_1)]\mathscr{A}_{\rm int}^{x_0}[x(\tau_2)]\rangle_{\Omega,c}^{x_0} + \cdots +$$
$$\left\{\prod_{j=1}^n\int_0^{\beta\hbar} d\tau_j\right\}\frac{(-1)^n}{n!\hbar^n}\left\langle\prod_{k=1}^n \mathscr{A}_{\rm int}^{x_0}[x(\tau_k)]\right\rangle_{\Omega,c}^{x_0} \qquad (9)$$

where $\mathscr{A}_{\rm int}^{x_0} = \mathscr{A} - \mathscr{A}_\Omega^{x_0}$ is the so-called inter-*action*, representing the perturbation to the harmonic reference state, and $\langle\cdots\rangle_{\Omega,c}^{x_0}$ is a cumulant which can be written in terms of expectation values $\langle\cdots\rangle_\Omega^{x_0}$ by the cumulant expansion, e.g.,

$$\langle\mathscr{A}_{\rm int}^{x_0}[x(\tau)]\rangle_{\Omega,c}^{x_0} \equiv \langle\mathscr{A}_{\rm int}^{x_0}[x(\tau)]\rangle_\Omega^{x_0} \qquad (10)$$

$$\langle\mathscr{A}_{\rm int}^{x_0}[x(\tau_1)]\mathscr{A}_{\rm int}^{x_0}[x(\tau_2)]\rangle_{\Omega,c}^{x_0} \equiv \langle\mathscr{A}_{\rm int}^{x_0}[x(\tau_1)]\mathscr{A}_{\rm int}^{x_0}[x(\tau_2)]\rangle_\Omega^{x_0} -$$
$$\{\langle\mathscr{A}_{\rm int}^{x_0}[x(\tau)]\rangle_\Omega^{x_0}\}^2 \qquad (11)$$

$$\langle\mathscr{A}_{\rm int}^{x_0}[x(\tau_1)]\mathscr{A}_{\rm int}^{x_0}[x(\tau_2)]\mathscr{A}_{\rm int}^{x_0}[x(\tau_3)]\rangle_{\Omega,c}^{x_0} \equiv$$
$$\langle\mathscr{A}_{\rm int}^{x_0}[x(\tau_1)]\mathscr{A}_{\rm int}^{x_0}[x(\tau_2)]\mathscr{A}_{\rm int}^{x_0}[x(\tau_3)]\rangle_\Omega^{x_0} -$$
$$3\langle\mathscr{A}_{\rm int}^{x_0}[x(\tau_1)]\mathscr{A}_{\rm int}^{x_0}[x(\tau_2)]\rangle_\Omega^{x_0}\langle\mathscr{A}_{\rm int}^{x_0}[x(\tau)]\rangle_\Omega^{x_0} +$$
$$2\{\langle\mathscr{A}_{\rm int}^{x_0}[x(\tau)]\rangle_\Omega^{x_0}\}^3 \text{ etc.} \qquad (12)$$

More importantly, Kleinert and co-workers derived a general expression for the expectation value of the form

$$\left\{\prod_{j=1}^n\int_0^{\beta\hbar} d\tau_j\right\}\left\langle\prod_{k=1}^n F_k[x(\tau_k)]\right\rangle_\Omega^{x_0}$$

in terms of Gaussian smearing convolution integrals[5,6]

$$\left\{\prod_{j=1}^n\int_0^{\beta\hbar} d\tau_j\right\}\left\langle\prod_{k=1}^n F_k[x(\tau_k)]\right\rangle_\Omega^{x_0} =$$
$$\left\{\prod_{j=1}^n\int_0^{\beta\hbar} d\tau_j\right\}\left\{\prod_{k=1}^n\int_{-\infty}^{\infty} dx_k\, F_k(x_k)\right\} \times$$
$$\frac{1}{\sqrt{(2\pi)^n\text{Det}[a_{\tau_k\tau_{k'}}^2(\Omega)]}}$$
$$\exp\left\{-\frac{1}{2}\sum_{\substack{k=1\\k'=1}}^n (x_k - x_0)a_{\tau_k\tau_{k'}}^{-2}(\Omega)(x_{k'} - x_0)\right\} \qquad (13)$$

where $\text{Det}[a_{\tau_k\tau_{k'}}^2(\Omega)]$ is the determinant of the $n \times n$-matrix consisting of the Gaussian width $a_{\tau_k\tau_{k'}}^2(\Omega)$, $a_{\tau_k\tau_{k'}}^{-2}(\Omega)$ is an element of the inverse matrix of $a_{\tau_k\tau_{k'}}^2(\Omega)$, and the Gaussian width is a function of the trial frequency Ω:

$$a_{\tau\tau'}^2(\Omega) = \frac{1}{\beta M\Omega^2}\left\{\frac{\beta\hbar\Omega}{2}\frac{\cosh[(|\tau - \tau'| - \beta\hbar/2)\Omega]}{\sinh(\beta\hbar\Omega/2)} - 1\right\} \qquad (14)$$

Using these smearing integrals in eq 13, the nth-order Kleinert variational perturbation (KPn) approximation, $W_n^\Omega(x_0)$, in eq 9 can be written in terms of ordinary integrations as follows[1]

$$W_n^\Omega(x_0) = -k_BT\ln Q_\Omega^{x_0} + \frac{k_BT}{\hbar}\int_0^{\beta\hbar} d\tau\langle V_{\rm int}^{x_0}[x(\tau_1)]\rangle_\Omega^{x_0} -$$
$$\frac{k_BT}{2!\hbar^2}\int_0^{\beta\hbar} d\tau_1\int_0^{\beta\hbar} d\tau_2\langle V_{\rm int}^{x_0}[x(\tau_1)]V_{\rm int}^{x_0}[x(\tau_2)]\rangle_{\Omega,c}^{x_0} + \cdots +$$
$$k_BT\frac{(-1)^{n+1}}{n!\hbar^n}\left\{\prod_{j=1}^n\int_0^{\beta\hbar} d\tau_j\right\}\left\langle\prod_{k=1}^n V_{\rm int}^{x_0}[x(\tau_k)]\right\rangle_{\Omega,c}^{x_0} \qquad (15)$$

where $V_{\mathrm{int}}^{x_0}[x(\tau)] = V[x(\tau)] - {}^1\!/_2 M\Omega^2[x(\tau) - x_0]^2$ (the kinetic energy terms in eq 4 and eq 5 cancel out).

As n tends to infinity, $W_n^{\Omega}(x_0)$ approaches the exact value of the centroid potential $W(x_0)$ in eq 1, which is independent of the trial Ω. But the truncated sum in eq 15 does depend on Ω, and the optimal choice of this trial frequency at a given order of KP expansion and at a particular centroid position x_0 (and temperature T) is determined by the least-dependence of $W_n^{x_0}(\Omega)$ on Ω itself. This is the so-called frequency of least dependence,[1] which provides a variational approach to determine the optimal value of Ω, $\Omega_{\mathrm{opt},n}(x_0)$.

Of particular interest is the special case when $n = 1$, which turns out to be identical to the original GTFK variational approach.[1,67−69] An important property of KP1 or the GTFK variational approach is that there is a definite upper bound for the computed $W_1^{\Omega}(x_0)$ by virtue of the Jensen-Peierls inequality, i.e., from eq 6 and eq 9

$$e^{-\beta W(x_0)} = Q_{\Omega}^{x_0}\left\langle \exp\left(-\frac{\mathcal{A} - \mathcal{A}_{\Omega}^{x_0}}{\hbar}\right)\right\rangle_{\Omega}^{x_0} \geq Q_{\Omega}^{x_0}\exp\left\langle -\frac{\mathcal{A} - \mathcal{A}_{\Omega}^{x_0}}{\hbar}\right\rangle_{\Omega}^{x_0} = e^{-\beta W_1^{\Omega}(x_0)} \tag{16}$$

Note that by choosing $\Omega = 0$ (i.e., the reference state for a free particle), KP1 or GTFK reduces to the Feynman-Hibbs approach.[8] For higher orders of n, unfortunately, it is not guaranteed that a minimum of $W_n^{x_0}(\Omega)$ actually exists as a function of Ω. In this case, the least dependent Ω is obtained from the condition that the next derivative of $W_n^{x_0}(\Omega)$ with respect to Ω is set to zero.[1,5,70] Consequently, Ω is considered as a variational parameter in the Kleinert perturbation theory such that $W_n^{x_0}[\Omega_{\mathrm{opt},n}(x_0)]$ is least-dependent on Ω.

This variational criterion relies on the uniformly and exponentially convergent property of the KP theory. Kleinert and co-workers proved that his theory exhibits this property in several strong anharmonic-coupling systems.[1,60,61] More importantly, this remarkably fast convergent property can also be observed even for computing the electronic ground-state energy of a hydrogen atom (3 degrees of freedom). The ground-state energy was determined by calculating the electronic centroid potential at the zero-temperature limit.[5] The accuracies of the first three orders of the KP theory for a hydrogen atom are 85%, 95%, and 98%, respectively.

In practice, for odd n, there is typically a minimum point in Ω,[1,5,70] but due to the alternating sign of the cumulants in eq 15, there is usually no minimum in Ω for even n. Nevertheless, the frequency of least-dependence for an even-order perturbation in n can be determined by locating the inflection point, i.e., the zero-value of the second derivative of $W_n^{x_0}(\Omega)$ with respect to Ω.[1,5,70] Since the KP expansion is uniformly and exponentially converged, Kleinert has demonstrated that the least-dependent plateau in $W_n^{x_0}(\Omega)$, which is characterized by a minimum point for odd n or by an inflection point for even n, grows larger and larger with increasing orders of n (e.g., Figure 5.16 in ref 1).

3. The Automated Integration-Free Path-Integral Method

A major obstacle in applying the KP theory to realistic molecular systems is the intricate n-dimensional space-time ($2n$ degrees of freedom) smearing integrals in eq 13 for the KPn expansion. The complexity of the smearing integrals increases considerably for multidimensional systems, where Ω becomes a $3N \times 3N$ matrix for N nuclei.[1,5] Thus, the KP theory quickly becomes numerically intractable beyond the first-order perturbation, i.e., the GTFK variational approach. To make the KP expansion feasible for many-body systems, we make use of the independent instantaneous normal mode (INM) approximation[27,43,96,97] to reduce the multidimensional potential to $3N$ one-body potentials. In the INM approximation, the total centroid effective potential for N nuclei is simplified as

$$W_n^{\Omega}(\{x_0\}^{3N}) \approx V(\{x_0\}^{3N}) + \sum_{i=1}^{3N} w_{i,n}^{\Omega}(q_i^{x_0}) \tag{17}$$

where $w_{i,n}^{\Omega}(q_i^{x_0})$ is the centroid potential for the INM coordinate $q_i^{x_0}$. Note that the INM coordinates are naturally decoupled through the second-order Taylor expansion. The INM approximation has also been used elsewhere.[27,43,96,97] This approximation should be especially suited for the KP expansion because the Gaussian convolution integrals in eq 13 exhibit the exponential decaying property from the centroid position.

For each INM, we further interpolate the potential energy along the INM coordinate in terms of an mth-order polynomial function because we have derived the analytical results of the path integrals in eq 15 up to 20th-order polynomials. Then, the optimal $\Omega_{\mathrm{opt}}(x_0)$ is numerically located by finding the least Ω-dependent centroid potential $W_n^{x_0}(\Omega)$ (section 2).[59] Hereafter, an mth-order polynomial representation of the original potential energy function obtained with an interpolating step size q Å both in the forward and backward directions along the normal mode coordinate at x_0 is denoted as P$m-q$A. Since the path integrals have been integrated analytically, the time-demanding Monte-Carlo or molecular dynamics samplings of eq 13 using different trial values of Ω to optimize the centroid potential is no longer necessary. Consequently, this essentially automated integration-free path integral (AIF-PI) approach is remarkably efficient and can be applied to chemical systems.[59] Analytical results were derived with Mathematica[98] and are available as Supplementary Material in ref 59.

In the INM approximation, previously we have shown that the computed quantum effects using the AIF-PI method are very encouraging for multidimensional systems, including systems that involve motions or vibrations of the lightest nucleus, hydrogen.[59] For example, we have computed the quantum partition function of a water molecule (3 degrees of freedom). At $T = 200$ K, the lowest temperature in which the exact partition function is available, the KP1 result is 77% of the exact result, while the KP2 value is 83% (note that the agreement in the corresponding free energy is much better, which are 99.2% and 99.4%, respectively, because of the logarithmic

Table 1. Classical and Quantum Canonical Partition Functions, and Free Energies of the Asymmetric Double-Well Potential at Various Temperatures[a]

T (K)	classical	accurate quantum	KP1	KP2	KP3	Mielke-Truhlar[b]	Doll-Myers[b]
			Canonical Partition Function				
1000	4.03E-01	3.12E-01	0.0	0.1	0.0	0.2	−0.2
500	1.85E-01	7.09E-02	−0.3	0.4	0.0	0.7	−0.2
400	1.47E-01	3.62E-02	−0.6	0.4	0.0	0.8	0.4
300	1.10E-01	1.19E-02	−1.4	0.3	0.0	0.9	2.7
200	7.28E-02	1.30E-03	−3.5	−0.5	−0.2	−0.6	10.0
100	3.63E-02	1.69E-06	−10.4	−2.4	−1.0	−8.2	50.6
50	1.81E-02	2.85E-12	−23.2	−6.7	−2.9	--	--
			Free Energy (kcal/mol)				
1000	1.808	2.314	0.0	−0.1	0.0	−0.2	0.2
500	1.676	2.629	0.1	−0.1	0.0	−0.3	0.1
400	1.524	2.638	0.2	−0.1	0.0	−0.2	−0.1
300	1.317	2.641	0.3	−0.1	0.0	−0.2	−0.6
200	1.041	2.641	0.5	0.1	0.0	0.1	−1.4
100	0.659	2.641	0.8	0.2	0.1	0.6	−3.1
50	0.399	2.641	1.0	0.3	0.1	--	--
0	0	2.641	1.2	0.4	0.2	--	--

[a] Signed percent errors (%) of different theoretical methods relative to the accurate quantum results are given. [b] Reference 27.

relationship between partition function and free energy), which is similar to the accuracy of the second-order Rayleigh−Schrödinger perturbation theory without resonance correction (86%).[59] Moreover, we have also computed the quantum correction factor for the collinear H_3 reaction (2 degrees of freedom), which is defined as the ratio of the quantum rate constant to the rate constant obtained by classical transition-state theory with quantum vibrational partition function but neglect tunneling effects.[59] In this reaction, both tunneling and vibrational quantum effects are important. At $T = 200$ K, again the lowest temperature in which the exact values is available, the KP1 and KP2 correction factors are 15 and 55, respectively, whereas the exact result is 46.

Although the INM approximation sacrifices some accuracy, in exchange, it makes possible analyses of specific contributions to the centroid potential W due to quantum mechanical vibration and tunneling. Positive and negative values of w_i in eq 17 raise (vibration) and lower (tunneling) the original potential V, respectively. In practice, real frequencies from the INM analysis often yield positive w_i's in eq 17, with dominant contributions due to zero-point effects (e.g., Sections 5A and 5B). By contrast, for imaginary frequencies in the INM, the values of w_i are often negative, corresponding to tunneling contributions (e.g., Sections 5A and 5C). To this end, we have also performed *ab initio* path-integral calculations[44−50] to determine kinetic isotope effects for a series of proton transfer reactions in water within the INM approximation. The agreement with experimental results is encouraging.[59]

Nevertheless, we are currently working on a formalism to systematically couple instantaneous normal modes in the AIF-PI method. We hope that one day this method could also be used by nonpath-integral experts or experimentalists as a 'black-box' for any given system.

4. Computational Details

The areas of the integrations for the classical configuration integrals involving $W(x_0)$ in eq 1 were chosen such that their values are converged at least up to 3 significant figures for all systems considered (Supporting Information), and the numerical integrations were performed with Mathematica.[98] To compute the centroid potential, all the original potentials were first mass-scaled in units of atomic mass unit (AMU). For the asymmetric double-well potential, we computed the eigenenergies by the Rayleigh−Ritz variational method[82−85] (Supporting Information), in which the Schrödinger equation was solved in a basis of 114 Gauss-Hermite polynomials. The partition functions calculated by summing over these eigenenergies with the Boltzmann factor are virtually identical to the quantum results reported in the Mielke-Truhlar paper.[27] Hence we treat them as the 'accurate quantum' values in Table 1 for the comparisons among all other approximate path-integral methods.

In the case of Morse potential, we have also numerically tested the sensitivity of using different orders of polynomial representation of the original potential and different interpolation steps, and we found that the results from both P10−0.02A and P20−0.08A are in excellent agreement. For example, at the equilibrium position of the original potential, their difference in the KP2 effective potential is less than 10^{-8} kcal/mol at a low temperature of 50 K. For the calculations of the Rayleigh−Ritz variational method and the Rayleigh−Schrödinger perturbation theory, we interpolated the Morse potential at the equilibrium position as P10−0.02A and P10−0.04A. Both interpolating polynomials return us the same energies at least up to 10^{-5} kcal/mol.

5. Results and Discussion

To illustrate the performance and accuracy of the Kleinert variational perturbation theory, we applied the AIF-PI method[59] to a number of well-studied systems, including an asymmetric double-well potential,[27] the Morse potential[80] corresponding to the bond vibrations of H_2,[99] HF,[100] and HCl,[99] and the Eckart potential[81] representing a model of proton transfer reactions.[100,101]

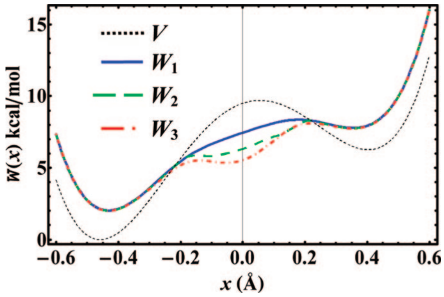


Figure 1. Computed (mass-scaled) centroid potential at the first- (W_1), second- (W_2), and third- (W_3) order Kleinert variational perturbation theory for an asymmetric double-well potential (V) at a temperature of 100 K.

A. Asymmetric Double-Well Potential. We first present the results for a particle of $M = 1224.259$ au (atomic units) in the one-dimensional asymmetric double-well potential

$$V(x) = b_4x^4 + b_2x^2 + b_1x + b_0 \tag{18}$$

where the four parameters have values of 0.01, −0.02, 0.005, and 0.01514754 au, respectively. This test case has been used by Mielke and Truhlar to validate their DPPI method based on a three-point trapezoidal approximation of the potential in the free-particle reference state.[27] Results based on the Doller-Myers approximation to the centroid potential[27] are also included for comparison (Table 1).

Key results are shown in Figure 1 which depicts the first three orders of KP effective potential as a function of the centroid variable x at 100 K, along with the original potential energy. It is of interest to notice that for all three perturbation levels, in the double-well regions, the path-integral centroid potential has a higher energy than the corresponding original potential energy primarily due to zero-point energy, whereas at the barrier region, the centroid potential is lowered in comparison with the original barrier, which may be attributed to tunneling effects. Furthermore, Figure 1 shows that in the double-well regions, the three perturbation results converge exceptionally well, but a noticeable progression of energy lowering effects is found at the barrier region. In fact, the original maximum point at $x \approx 0$ Å, in which the frequency for the INM is imaginary, has become a local minimum at the KP3 level purely due to quantum tunneling effects. Note that for this one-dimensional case, the centroid potential is equivalent to the centroid potential of mean force along the position coordinate. Thus, if the potential of mean force is used in path-integral quantum transition state theory (PI-QTST),[11,13,69,101,102] it seems that the lowest perturbation level at KP1 may underestimate tunneling and could lead to noticeable errors in rate calculations.

The quantum partition function and the corresponding free energy from the Kleinert variational perturbation theory at different temperatures are summarized in Table 1, along with the results from the earlier studies.[27] Note that the dominant contribution to the partition function in the configuration integral is from the centroid potential near the global minimum (in which the INM frequencies are all real). This is reflected in the good agreement among all three perturbation levels (Figure 1 and Table 1). Furthermore, the absolute values of the computed partition function have greater errors in comparison with the accurate results than the corresponding free energies due to the exponential relationship between the two quantities. Consequently, the computed free energies uniformly have smaller errors than the absolute partition functions for all approximate methods. In the entire temperature range that has been considered, the KP theory, particularly at the KP2 and KP3 levels, consistently yields the most accurate results among all methods listed in Table 1. The DDPI method slightly outperforms the GTFK variational approach which is equivalent to KP1, but there are significant errors at low temperatures. In this case, the second- and third-order perturbation results are in excellent agreement (99%) with accurate quantum results. At 50 K, the KP1 value deteriorates significantly, whereas the results obtained using KP2 and KP3 are still in good accord with the accurate results. Interestingly, Mielke and Truhlar pointed out that although the GTFK variational approach (KP1) is the most accurate method that they have considered, the method is too expensive for computing molecular partition functions.[27] In the present AIF-PI method, the path integrals are determined analytically, which makes the daunting task of path-integral simulation a trivial problem, allowing the variational frequencies to be optimized quickly and efficiently.

An important property of the centroid potential is that at the limit of zero-temperature the energy and position of the global minimum correspond to, respectively, the ground-state energy and the expectation value of position in the ground-state determined by wave mechanics[1,14,15,68]

$$\lim_{T\to 0} W_T(x_{\min}) = E_0 \tag{19}$$

and

$$x_{\min} = \langle \psi_0 | x | \psi_0 \rangle \tag{20}$$

where x is the position operator, and $x_{\min}$ and $W_T(x_{\min})$ are, respectively, the coordinate and value at the global minimum of the centroid potential. In eq 19 and eq 20, ψ_0 is the nuclear ground-state wave function and E_0 is the lowest eigenenvalue of the Hamiltonian, i.e., the zero-point energy. We have derived the closed-form expressions (Supporting Information) for these quantities up to the KP1/P20, KP2/P20, and KP3/P6 levels of theory, which can be evaluated at no additional computational costs once the centroid potential is optimized. In contrast, it would be extremely difficult to obtain converged results at 0 K using Monte Carlo or molecular dynamics path integral simulations. In the KP theory, the quantum ground-state energy and the expectation value of particle position can thus be obtained simply by performing centroid potential energy minimizations.

Listed in Table 2 are the calculated ground-state energy and the expectation value of position for the asymmetric double-well potential from the KP theory. The estimated zero-point energies are 2.672, 2.650, and 2.645 kcal/mol at the KP1, KP2, and KP3 level of theory, respectively, which represents errors of 1.18%, 0.35%, and 0.16% from the exact value of 2.641 kcal/mol. The expectation value of the particle position is shifted away from the coordinate at the minimum of the original potential due to asymmetry of the potential. The corresponding KP results are in remarkable agreement

Table 2. Classical and Quantum Ground-State Energy (kcal/mol), and the Expectation Value of the Particle Position (Å) in an Asymmetric Double-Well Potential along with the Minimum Energy of the Centroid Potential and Position from the First Three-Order of the Kleinert Variational Perturbation Theory

ground state	classical	accurate quantum	KP1	KP2	KP3
energy	0	2.641	2.672	2.650	2.645
$\langle x \rangle$	−0.55958 ($x_{\min}$)	−0.52189	−0.52329	−0.52272	−0.52231

Table 3. Parameters of the Morse Potential for Hydrogen Chloride, Hydrogen Fluoride, and Hydrogen Molecules

molecule	D_e (kcal/mol)	r_0 (Å)	γ (Å)
HCl[a]	106.48594	1.274577	0.535536
HF[b]	136.30000	0.916600	0.452243
H_2[a]	109.48232	0.741589	0.514992

[a] Reference 99. [b] Reference 100.

with the accurate result, having errors only of 0.27%, 0.16%, and 0.08% in comparison with the exact value of −0.52189 Å.

B. The Morse Potential for Bond Vibration. The Morse potential[80] is selected to model the bond vibrations of three diatomic molecules: H_2, HF, and HCl[99,100]

$$V(r) = D_e\left[1 - \exp\left(-\frac{r - r_0}{\gamma}\right)\right]^2 \tag{21}$$

where r is the bond length, r_0 is the equilibrium distance, D_e is the bond dissociation energy, and γ is a parameter related to the harmonic frequency by $\omega_0 = (1/\gamma)\sqrt{(2D_e/\mu)}$ with μ being the reduced mass of a diatomic molecule. The parameters for the three molecules are given in Table 3.[99,100]

We first examine the temperature dependence of the quantum partition function Q_{qm} from the bound states for hydrogen fluoride using the KP theory. The computational results obtained by using a tenth-order polynomial with 0.02 Å interpolation steps (P10−0.02A) for the Morse potential at the KP1 and KP2 levels of theory are listed in Table 4, along with the exact results for a temperature ranging from 50 to 1000 K. Good agreement is found between results from KP1 or KP2 calculations and the exact values. Even at 50 K, the KP1 partition function is still within 75% of the exact value (−25% error), while the KP2 theory shows a remarkable 96.4% accuracy (−3.6% error). As noted above, although the absolute partition function may show greater errors due to the exponential dependence on the centroid potential, the computed free energies are less sensitive as illustrated in Table 4. For example, at the zero-temperature limit, the computed free energies using KP1 and KP2 are within 0.52% and 0.07% of the exact zero-point energy, respectively.

In Tables 5 and 6, we summarize the computed zero-point energies and bond length expectation values at the global minima of the centroid potentials from both KP1 and KP2 levels of theory, along with the exact results. In all cases, the agreement with the exact quantum data is excellent. For H_2, which is expected to have the largest quantum effects due to its small reduced mass, the error in the computed ground-state energy (lowest value in the centroid potential) is only 0.15% at the KP2 level of theory. All calculated expectation values of the bond length both at KP1 and KP2 levels are within 0.0001 Å of the exact quantum results, which are two-order magnitude more accurate than the classical equilibrium positions.

Figure 2 illustrates the KP1 and KP2 centroid potentials at 200 K. It is interesting to note that approximately before the inflection point of the original potential (in which the INM frequencies are real), the computed centroid potentials are above the original Morse potential, dominated by zero-point vibrational effects, whereas approximately beyond the inflection point (in which the INM frequencies are imaginary), the centroid potentials are below the Morse potential.

To shed light on the relationship between KP theory and the use of the harmonic frequency of the Morse potential in approximating quantum energy as well as traditional (wave function) perturbation and variational theories, we have determined the harmonic zero-point energy (i.e., $\hbar\omega_0/2$), and the ground-state energy using the Rayleigh−Ritz variational approach,[82−85] and the second-order Rayleigh−Schrödinger perturbation theory.[63,86,87] All results are also listed in Table 5. In the Rayleigh−Ritz method, the ground-state harmonic eigenfunction centered at r_0 is variationally optimized by adjusting the Gaussian width a:

$$\tilde{\varphi}(r) = \frac{1}{(2\pi a^2)^{1/4}}\exp\left[-\frac{(r - r_0)^2}{4a^2}\right] \tag{22}$$

The Rayleigh−Schrödinger perturbation was carried out up to the second order, in which the excited states are constructed by using the angular frequency $\tilde{\Omega}$ deduced from the Rayleigh−Ritz optimization

$$E^{(2)} = \frac{\hbar\tilde{\Omega}}{2} + \langle\tilde{\varphi}_0|V(r) - \frac{1}{2}\mu\tilde{\Omega}^2(r - r_0)^2|\tilde{\varphi}_0\rangle + \sum_{k\neq 0}\frac{\left|\left\langle\tilde{\varphi}_k|V(r) - \frac{1}{2}\mu\tilde{\Omega}^2(r - r_0)^2|\tilde{\varphi}_0\right\rangle\right|^2}{\mathscr{E}_k - \mathscr{E}_0} \tag{23}$$

where $\tilde{\varphi}_0$ is the wave function in eq 22 but with the optimized Gaussian width a being optimized, $\tilde{\Omega} = \hbar/2\mu a^2$, $\tilde{\varphi}_k$ are the eigenfunctions for the harmonic system with the angular frequency $\tilde{\Omega}$, and $\mathscr{E}_k$ are the eigenenergies $\hbar\tilde{\Omega}(k + {}^1/_2)$.

In Table 5, the zero-point energies computed using pure harmonic frequencies have errors greater than results both from the KP1 and KP2 theory, suggesting anharmonicity is indeed important for high frequency vibrations involving hydrogen atoms. Surprisingly, the Rayleigh−Ritz variational approach performs worse than the harmonic approximation, and this may be attributed to the fact that the location of the wave function is fixed at the minimum of the Morse potential energy function. If the center of the Gaussian wave function in eq 22 is also treated as a variational parameter along with the width, the Rayleigh−Ritz variational results reduces to the KP1 value at the zero-temperature limit. Therefore, the variationally opti-

Table 4. Classical and Quantum Canonical Partition Functions, and Free Energies of the Morse Potential for Hydrogen Fluoride at Various Temperatures[a]

T (K)	classical	quantum	KP1/P10−0.02A	KP2/P10−0.02A
		Canonical Bound Partition Function		
1000	1.73E-01	5.61E-02	−0.2	0.0
500	8.61E-02	3.12E-03	−1.0	0.0
400	6.88E-02	7.38E-04	−1.8	0.0
300	5.16E-02	6.67E-05	−3.0	−0.1
200	3.44E-02	5.45E-07	−5.1	−0.4
100	1.72E-02	2.97E-13	−12.1	−1.3
50	8.58E-03	8.84E-26	−24.9	−3.6
		Free Energy (kcal/mol)		
1000	3.486	5.724	0.1	0.0
500	2.436	5.733	0.2	0.0
400	2.127	5.732	0.2	0.0
300	1.767	5.732	0.3	0.0
200	1.340	5.732	0.4	0.0
100	0.807	5.732	0.4	0.0
50	0.473	5.732	0.5	0.1
0	0	5.732	0.5	0.1

[a] Signed percent errors (%) of different theoretical methods relative to the accurate quantum results are given.

Table 5. Computed Ground State Energies (kcal/mol) for Hydrogen Chloride, Hydrogen Fluoride, and Hydrogen Molecules from the Morse Potential Using the Harmonic Approximation, Rayleigh−Ritz (RR) Variational Approach, Second-Order Rayleigh−Schrödinger Perturbation Theory (RS2), and First and Second Orders of the Kleinert Variational Perturbation Theory (KP1 and KP2)

molecule	quantum	harmonic	RR	RS2	KP1/P10−0.02A	KP2/P10−0.02A
HCl	4.231	4.274	4.348	4.238	4.253	4.234
HF	5.732	5.793	5.899	5.742	5.762	5.736
H_2	6.193	6.284	6.437	6.213	6.238	6.202

Table 6. Computed Expectation Value of Particle Position and the Minimum Centroid Potential Coordinate (Å) for H_2, HF, and HCl Using KP1 and KP2 Theory[a]

molecule	classical	quantum	KP1/P10−0.02A	KP2/P10−0.02A
HCl	1.274577	1.29094	1.29086	1.29089
HF	0.916600	0.93124	0.93117	0.93121
H_2	0.741589	0.76423	0.76409	0.76416

[a] The equilibrium bond distances of the original (classical) potential are also given.

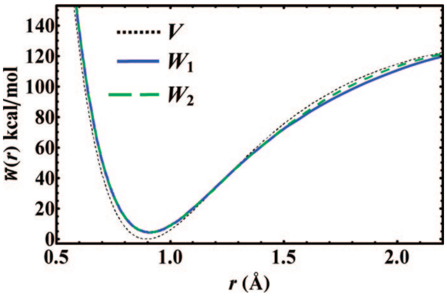


Figure 2. The first- and second-order (mass-scaled) centroid potentials (W_1 and W_2) from the KP theory compared with the Morse potential (V) for hydrogen fluoride as a function of the centroid bond length at 200 K.

mized position of the Rayleigh−Ritz wave function may be interpreted as the centroid position in path integrals.[1,67−69] The deficiency without optimizing the location of the trial wave function is partially recovered by the second-order Rayleigh−Schrödinger perturbation theory (Table 5), which has an accuracy between KP1 and KP2. Although it is tempting to optimize the center of the wave functions $\tilde{\varphi}_k$ used in Rayleigh−Schrödinger perturbation theory, this is not possible because the perturbation theory is not a variational method. Nevertheless, for a symmetric potential, Kleinert showed that at the zero-temperature limit, the KP theory is identical to the Rayleigh−Schrödinger perturbation theory, provided that the global minimum point in the centroid potential is chosen as the center of the wave function.[1] In general, Kleinert's variational perturbation theory resembles the combined features of both the Rayleigh–Ritz variational method and the Rayleigh–Schrödinger perturbation theory in wave mechanics.

C. Symmetric and Asymmetric Eckart Potentials. The Eckart potential[81,103] (Supporting Information) is a popular model for testing a rate theory for chemical reactions because the quantum result is known exactly. Since the reactant and product states are unbound free particles, which have identical classical and quantum partition functions, the difference between classical and quantum rate constants for the Eckart potential is entirely due to tunneling. Often, quantum tunneling effect[88,89] is expressed as the ratio of the quantum rate constant to that of classical transition state theory[104] (TST)

$$\kappa = \frac{k_{qm}}{k_{TST}} = \beta e^{\beta V_{\max}} \int_0^\infty \gamma(E) e^{-\beta E} dE \qquad (24)$$

where k_{qm} and k_{TST} are the quantum and TST rate constants, and $\gamma(E)$ is the transmission probability at energy E, which can be determined exactly for the Eckart potential (Supporting Information), and $V_{\max}$ is the barrier height. κ is called the quantum tunneling correction factor or transmission coefficient.[103] In path-integral quantum transition state theory

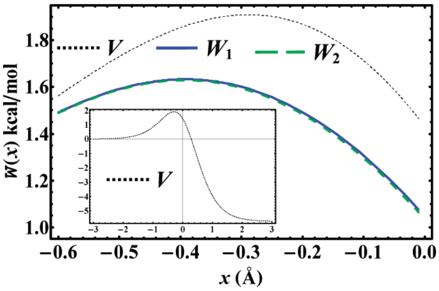


Figure 3. Comparison of the mass-scaled KP1 (W_1) and KP2 (W_2) centroid potentials with the corresponding Eckart potential (V) near the barrier top at 82 K.

(PI-QTST),[11,13,69,101,102] κ is approximated as follows (when there is no correction for recrossings)

$$\kappa \approx \exp[-\beta(W_{\max} - V_{\max})] \tag{25}$$

where $W_{\max}$ is the maximum energy of the centroid potential at the PI-QTST transition state. Note that $W_{\max}$ is not necessarily located at the same position as $V_{\max}$ (e.g., Figure 3 and Table 10), although all $W_{\max}$ we found are in the region where the INM frequencies are imaginary. In this study, we consider both the symmetric and asymmetric situations and compare the KP results with those from previous studies.

For the symmetric Eckart potential, we used a set of parameters corresponding to a protium tunneling through a barrier of $V_{\max} \approx 5.7$ kcal/mol with an angular frequency of $\omega^* = 1047.2$ cm^{-1} at the top of the barrier (Supporting Information). This set of parameters has been widely used both in analytic theories and in path integral Monte Carlo (PIMC) simulations.[40,69,103] Table 7 summarizes the computed quantum correction factor κ using the present KP1 and KP2 theory, along with results obtained from a diagrammatic approach by Cao and Voth (CV),[40] and from PIMC simulations.[69] Due to symmetry, $W_{\max}$ is located at the same position as $V_{\max}$. The theoretical approach used in the Voth-Chandler-Miller paper[69] (VCM) yields identical results as that from KP1 or the GTFK variational approach, whereas the diagrammatic method of Cao and Voth is similar to KP2 without variational optimization of the angular frequency. In our study, we have used a 20th-order polynomial (P20−0.2A) representation of the Eckart potential, fitted in the region of $x_0 \pm 2$ Å. Using this interpolated polynomial potential, the KP1 results are nearly identical to those obtained by Voth at al. without using potential interpolation.[69] Hence, the computational accuracy by P20−0.2A representation is clearly reasonable in the present calculations.

In comparison with the exact results, the KP1 theory shows noticeable deviations at low temperature, while the results obtained using the KP2 theory, the CV approach, and the PIMC simulation are all very accurate even at a temperature as low as 126 K (Table 7).

Moreover, we list the computed kinetic isotope effects[89,105] (KIE) for protium and deuterium transfer reactions at the KP1 and KP2 levels in Table 8, which are compared with values obtained previously with PIMC simulations (Supporting Information).[100] A similar trend is observed in that although only the KP2 theory is accurate at lower temperatures, at room temperature, and above, both KP1 and KP2 perform very well in comparison with PIMC simulations and the exact data.

We now turn our attention to the asymmetric Eckart potential (Supporting Information) which has been used by Jang et al.[101] ($\omega^* \approx 340$ cm^{-1}) to test the PI-QTST.[11,13,69,101,102] Table 9 shows that good accord is obtained between the PI-QTST quantum correction factor from path-integral molecular dynamics (PIMD) simulations and the present perturbation result, particularly at the KP2 level of theory. It is interesting to notice that at low temperature, both the PI-QTST simulation and the KP theory overestimate the tunneling effect (Table 9), whereas it is underestimated for the symmetric potential (Table 7). We attribute the difference in the asymmetric Eckart potential to the inclusion of energy E smaller than the reaction energy A in calculating the centroid potential, though there is no contribution to tunneling transmission $\gamma(E)$ for energy less than A (Supporting Information). This discrepancy may be resolved by integrating only the closed paths in which the quantum-statistical action $\mathscr{A}[x(\tau)]$ is larger than or equal to A in the centroid potential calculations:

$$W(x_0) = -k_B T \ln\left[\sqrt{\frac{2\pi\hbar^2}{Mk_BT}} \oint \mathscr{D}[x(\tau)]\delta(\bar{x} - x_0)|_{\mathscr{A}\geq A} \times \exp\{-\mathscr{A}[x(\tau)]/\hbar\}\right] \tag{26}$$

Jang et al.[101] proposed a way to alleviate this problem by shifting the lower energy region of the asymmetric potential to match the high energy asymptotic value, i.e., by effectively using a less asymmetric potential.

In Table 10, we report the $x_{\max}$ values, at which the centroid potential is at the maximum, at different temperatures. The $x_{\max}$ values correspond to the PI-QTST saddle point (or transition state), which is shifted away from the position of the original barrier. At 61 K, $x_{\max}$ deviates from the classical transition state by more than 0.2 Å.

D. Optimization of the Variational Frequency. Both the nth-order centroid potential W_n^Ω and the associated optimal frequency $\Omega_{\mathrm{opt},n}$ are functions of the centroid position x_0 and temperature T. As n tends to infinity, $W_n^{x_0,T}$ (Ω) becomes independent of Ω, and $W_n^{x_0,T}$ (Ω) is exact. This, in fact, provides a variational procedure for determining $W_n^{x_0,T}$ (Ω) on the basis of least dependence on Ω (Section 2).[1] Thus, at a given position x_0 and temperature T, we have the centroid potential $W_n^{x_0,T}$ (Ω) as a function of Ω. The optimal value $\Omega_{\mathrm{opt},n}(x_0,T)$ is at the $W_n^{x_0,T}$ (Ω) minimum or is located at the point that $W_n^{x_0,T}$ (Ω) has the least Ω-dependence if a minimum does not exist, i.e., the second derivative is zero. The latter corresponds to an inflection point (Section 2). In this section, we examine some features of the optimal value of the variational parameter $\Omega_{\mathrm{opt},n}(x_0,T)$ at different perturbation levels and the dependence of the centroid potential on Ω at different temperatures. All discussions are based on the asymmetric double-well potential discussed in Section 5A.

Figure 4 shows the square of the optimal variational frequencies $\Omega_{\mathrm{opt},n}(x_0,T)$, where $n = 1$, 2, and 3, along the

Table 7. Computed Quantum Transmission Coefficient κ for the Symmetric Eckart Barrier at Various Temperatures[a,b]

		κ					
$\beta\hbar\omega^*$	T (K)	exact	KP1/P20−0.2A	KP2/P20−0.2A	VCM	Cao-Voth	PIMC
2	753	1.224	1.169	1.169	1.2	--	1.2
3	502	1.525	1.419	1.420	1.4	--	1.4
4	377	2.071	1.870	1.872	1.9	--	1.9
5	301	3.102	2.696	2.708	2.7	--	2.7
6	251	5.199	4.283	4.353	4.4	4.4	4.4
8	188	21.77	14.71	16.39	15.0	17.0	17.0
10	151	161.9	74.29	112.0	73.0	110.6	105.0
12	126	1973	514.6	1484	514.0	1278.0	1240.0

[a] Reference 69. [b] Reference 40.

Table 8. Kinetic Isotope Effects (KIE) on the Protium and Deuterium Transfer over the Symmetric Eckart Potential at Various Temperatures

	KIE (protium/deuterium)			
T (K)	exact	KP1/P20−0.2A	KP2/P20−0.2A	PIMC[a]
500	1.232	1.186	1.186	1.19
400	1.374	1.306	1.307	1.31
350	1.511	1.421	1.423	1.43
300	1.756	1.623	1.630	1.62
250	2.283	2.036	2.068	2.08
200	3.840	3.110	3.328	3.33
150	12.17	7.206	10.29	11.10[b]

[a] Reference 100. [b] Computed in this work using the same PIMC program in ref 100.

Table 9. Transmission Coefficient κ for the Asymmetric Eckart Barrier at Various Temperatures

		κ			
$\beta\hbar\omega^*$	T (K)	exact	KP1/P20−0.2A	KP2/P20−0.2A	PIMD[a]
2	245	1.195	1.178	1.178	1.17
4	122	2.019	1.985	1.989	1.97
6	82	5.387	5.528	5.668	5.69
8	61	27.27	31.55	35.39	36.6

[a] Reference 101.

Table 10. Temperature Dependence of $x_{\max}$ of the Asymmetric Eckart Barrier

		$x_{\max}$ (Å)		
$\beta\hbar\omega^*$	T (K)	classical	KP1/P20−0.2A	KP2/P20−0.2A
2	245	−0.28645	−0.30814	−0.30813
4	122	−0.28645	−0.33691	−0.33687
6	82	−0.28645	−0.39049	−0.39212
8	61	−0.28645	−0.48334	−0.49329

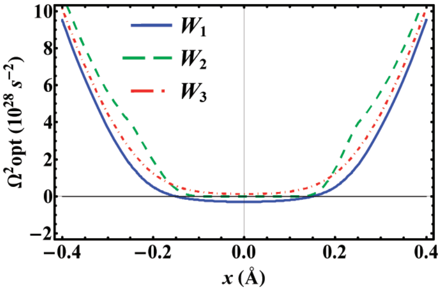


Figure 4. The $\Omega^2_{\mathrm{opt},1}(x)$, $\Omega^2_{\mathrm{opt},2}(x)$, and $\Omega^2_{\mathrm{opt},3}(x)$ for the mass-scaled asymmetric double-well potential at $T = 100$ K.

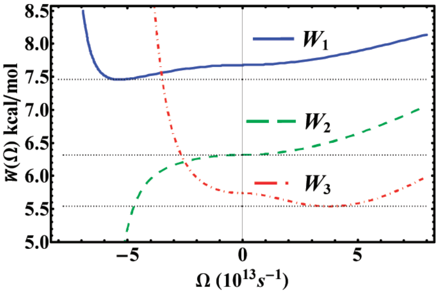


Figure 5. $W_1^{x_0,T}$ (Ω), $W_2^{x_0,T}$ (Ω), and $W_3^{x_0,T}$ (Ω) at $x_0 = 0$ and $T = 100$ K for the asymmetric double-well potential. Note that the imaginary axis of Ω is represented by the negative axis.

coordinate position at 100 K, which are used to determine the corresponding centroid potentials at the three perturbation levels. First, $\Omega_{\mathrm{opt},n}(x_0,T)$ can be either a real or an imaginary quantity, depending on the potential surface, the coordinate position, and temperature. For the KP1 theory, the variational frequency is imaginary in the barrier region of the double-well potential, whereas $\Omega_{\mathrm{opt},3}(x_0,100\text{ K})$ has real values throughout. Interestingly, for an extended range around $x_0 = 0$, $\Omega_{\mathrm{opt},2}(x_0,100\text{ K})$ has an optimal value of zero. In Figure 5, we focus on a given centroid position at $x_0 = 0$ and $T = 100$ K and illustrate the dependence of the three KP centroid potentials on the variational frequency Ω. In this figure, the imaginary axis of Ω is represented by the negative axis (Note: $W_n^{x_0,T}$ (Ω) is an even function of Ω.). In this case, the optimal frequency is imaginary for W_1 with a value of $5.35694 \times 10^{13} i\ \mathrm{s}^{-1}$, real for W_3 with a value of $3.78053 \times 10^{13}\ \mathrm{s}^{-1}$, and zero s^{-1} for W_2. For reference, the INM frequency at $x_0 = 0$ is imaginary, with a value of $2.36308 \times 10^{14} i\ \mathrm{s}^{-1}$.

Figure 6 depicts the Ω-dependence of $W_n^{x_0,T}$ (Ω) at $x_0 = -0.55958$ Å (in which the original potential [eq 18] is at the global minimum) for three different temperatures: $T = 500$, 386, and 100 K. Under these conditions, the optimal values of the variational frequency are all real, and we see that the change in the centroid potential becomes less sensitive at a higher order perturbation when Ω is large (i.e., greater than the optimal value). Furthermore, the KP2 centroid potential exhibits an inflection point rather than a minimum as a function of Ω in all three temperatures because of the alternating signs in the cumulant expansion for even-order terms in the KP theory (Section 2).[1,5,70] These observations are consistent with those described by Kleinert in ref 1.

In contrast to the large value of Ω, when Ω is small, at lower temperatures the centroid potential becomes more

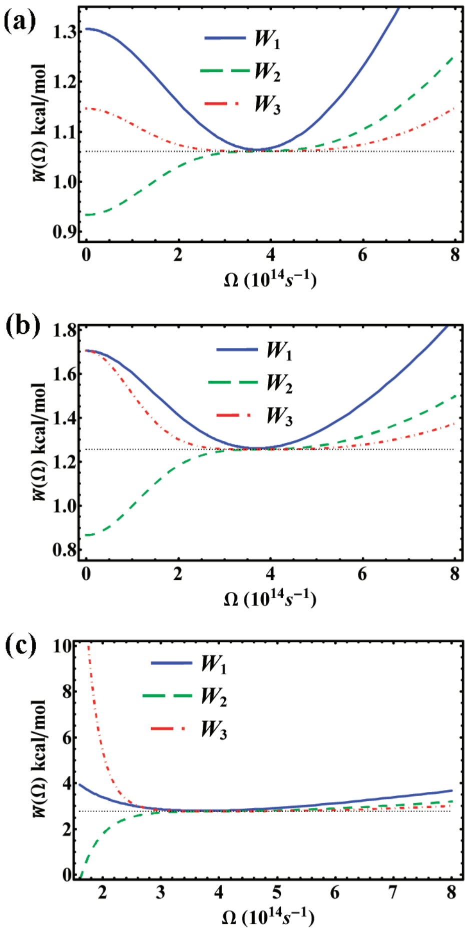


Figure 6. $W_1^{x_0,T}$ (Ω), $W_2^{x_0,T}$ (Ω), and $W_3^{x_0,T}$ (Ω) at $x_0 = -0.55958$ Å and (a) $T = 500$ K, (b) $T = 385.96$ K, and (c) $T = 200$ K for the asymmetric double-well potential.

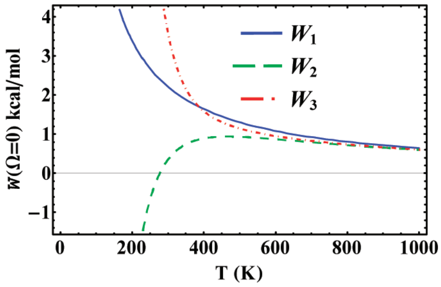


Figure 7. Temperature-dependence of $W_1^{x_0}(\Omega = 0)$, $W_2^{x_0}(\Omega = 0)$, and $W_3^{x_0}(\Omega = 0)$ at $x_0 = -0.55958$ Å for the asymmetric double-well potential.

sensitive at higher order perturbations (Figure 6). At the limit of zero-Ω (analytical expressions for zero-Ω limit are available in the Supporting Information), the centroid potential corresponds to that using a free-particle reference frame, which is the framework used in the Feynman-Hibbs approach.[8] In this limit, the perturbative inter-action in the KP theory [i.e., $V_{\rm int}^{x_0}$ in eq 15 or $\mathcal{A}_{\rm int}^{x_0}$ in eq 9] is the original potential itself. In Figure 7, the centroid potential at zero-Ω limit is shown as a function of temperature at $x_0 = -0.55958$ Å for the first three orders of KP theory. W_1 and W_3 coincide at temperature of 386 K. Although the results converge in the high-temperature limit, they diverge at lower temperatures. This behavior suggests that (not surprisingly) the free-particle reference state is not particularly a good choice in a general perturbation expansion.

6. Concluding Remarks

The Kleinert variational perturbation theory is a systematic, fast convergent method for treating internuclear quantum effects in molecular systems. In the KP theory, the angular frequency Ω for a harmonic reference state is variationally optimized at a given centroid position x_0 and temperature T, and the exact quantum partition function is obtained by systematically incorporating anharmonic corrections to the centroid potential of this reference system. Despite the numerous attractive features of the KP theory, it has not been used in chemical applications beyond the first-order perturbation, i.e., the so-called Giachetti-Tognetti-Feynman-Kleinert (GTFK) variational approach. The primary drawback is the need to optimize the variational frequency, which becomes an $n \times n$ matrix, by path-integral calculations. The coupled path-integral effective potential optimization and evaluation of Gaussian smearing functions in the KP theory is a daunting computational task, and only until very recently, a practical procedure has been devised for condensed phase simulations at the KP1 level of theory.[77,78] Making use of the instantaneous normal mode approximation, which reduces a system of $3N$ degrees of freedom (where N is the number of particles) to $3N$ one-dimensional problems, we have developed an analytical method to obtain the centroid potential as a function of the variational parameter in the KP theory,[59] avoiding numerical path-integral Monte Carlo or molecular dynamics simulations, especially at the zero-temperature limit. Consequently, the variational procedure in the KP theory can be efficiently carried out, and, thus, higher order perturbations can be performed for realistic chemical applications. Previously, we have demonstrated that in the INM approximation, the AIF-PI method is still accurate for computing the quantum partition function of a water molecule (3 degrees of freedom) and the quantum correction factor for the reaction rate of the collinear H_3 reaction (2 degrees of freedom).[59]

In the present study, we further test the accuracy and properties of KP theory by using the first three-order perturbations to determine the zero-point energy, quantum partition function, and tunneling factor for systems including an asymmetric double-well potential, the bond vibrations of H_2, HF, and HCl represented by the Morse potential, and a hydrogen-transfer barrier modeled by the Eckart potential. The following general conclusions are drawn from these calculations:

(1) Kleinert's variational perturbation theory is an extremely accurate method for treating internuclear quantum-statistical effects and for obtaining path-integral centroid potentials. Although the lowest (first-order) level perturbation theory, KP1, which is identical to the GTFK variational method, shows noticeable deviations (by more than 25%) from the exact quantum results (including the partition function and tunneling factor), at temperatures below 100 K, the second- and third-order perturbations, KP2 and KP3, remain accurate, within 96% of the exact values for all systems considered.

(2) Using our newly derived analytical results (Supporting Information), the minimum value of the centroid potential at the zero-temperature limit is in excellent agreement with the ground-state energy (zero-point energy), and the position of the centroid-potential minimum is the expectation value of particle position in wave mechanics.

(3) In comparison with the Rayleigh−Ritz (RR) variational approach and the Rayleigh−Schrödinger (RS) perturbation theory in wave mechanics, the results from the KP theory in path-integral quantum mechanics combine both features of RR variation (optimization of both the center and width of wave function) and RS perturbation that includes dynamic correlations. Consequently, the Kleinert perturbation theory converges exceedingly fast, and the KP2 level of theory can yield accurate results for computing kinetic isotope effects in chemical reactions.[59]

(4) Finally, the centroid potential obtained from the Kleinert perturbation calculations can be used in combination with path-integral quantum transition state theory (PI-QTST) to estimate the rate constant of chemical reactions and can be applied to condensed phase systems and enzymatic processes.

Acknowledgment. We thank Dr. Timothy J. Giese for the program to compute the eigenenergies of the asymmetric double-well potential and Prof. Donald G. Truhlar for discussions on an early draft of this manuscript. This work is partially supported by the National Institutes of Health under grant number GM46736 and by the Office of Naval Research for the development of integrated tools for studying chemical kinetics.

Supporting Information Available: Complete computational details; eigenenergies of the asymmetric double-well potential; and instructions to obtain analytical closed forms of KP1/P20, KP2/P20, and KP3/P6 at the zero-Ω and zero-temperature limits in the formats of Mathematica notebook and FORTRAN. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Kleinert, H. *Path integrals in quantum mechanics, statistics, polymer physics, and financial markets*, 3rd ed.; World Scientific: Singapore, River Edge, NJ, 2004; p xxvi, 1468 p. For the variational perturbation theory, see chapters 3 and 5.

(2) Kleinert, H. *Phys. Lett. A* **1993**, *173*, 332–342.

(3) Jaenicke, J.; Kleinert, H. *Phys. Lett. A* **1993**, *176*, 409–414.

(4) Kleinert, H.; Meyer, H. *Phys. Lett. A* **1994**, *184*, 319–27.

(5) Kleinert, H.; Kürzinger, W.; Pelster, A. *J. Phys. A: Math. Gen.* **1998**, *31*, 8307–8321.

(6) Bachmann, M.; Kleinert, H.; Pelster, A. *Phys. Rev. A* **1999**, *60*, 3429–3443.

(7) Janke, W.; Pelster, A.; Schmidt, H.-J.; Bachmann, M. *Fluctuating paths and fields: festschrift dedicated to Hagen Kleinert on the occasion of his 60th birthday*; World Scientific: River Edge, NJ, 2001; p xxi, 850 p. For the variational perturbation theory, see part III.

(8) Feynman, R. P.; Hibbs, A. R. *Quantum mechanics and path integrals*; McGraw-Hill: New York, NY, 1965; p xiv, 365 p. For the applications in quantum statistics, see chapters 10 and 11. Corrections to the errata in the book: http://www.oberlin.edu/physics/dstyer/FeynmanHibbs/ and http://www.physik.fu-berlin.de/∼kleinert/Feynman-Hibbs/.

(9) Gillan, M. J. *Phys. Rev. Lett.* **1987**, *58*, 563–6.

(10) Gillan, M. J. *J. Phys. C: Solid State Phys.* **1987**, *20*, 3621–3641.

(11) Voth, G. A. *J. Phys. Chem.* **1993**, *97*, 8365–8377.

(12) Cao, J.; Voth, G. A. *J. Chem. Phys.* **1994**, *101*, 6168–83.

(13) Voth, G. A. *Adv. Chem. Phys.* **1996**, *93*, 135–218.

(14) (a) Ramírez, R.; López-Ciudad, T.; Noya, J. C. *Phys. Rev. Lett.* **1998**, *81*, 3303–3306. (b) Comment: Andronico, G.; Branchina, V.; Zappala, D. *Phys. Rev. Lett.* **2002**, *88*, 178901. (c) Reply to Comment: Ramirez, R.; López-Ciudad, T. *Phys. Rev. Lett.* **2002**, *88*, 178902.

(15) Ramírez, R.; López-Ciudad, T. *J. Chem. Phys.* **1999**, *111*, 3339–3348.

(16) Feynman, R. P. *Statistical mechanics; a set of lectures*; W. A. Benjamin: Reading, MA, 1972; p xii, 354 p.

(17) Brown, L. M. *Feynman's thesis: a new approach to quantum theory*; World Scientific: Singapore, Hackensack, NJ, 2005; p xxii, 119 p.

(18) Feynman, R. P. *Rev. Mod. Phys.* **1948**, *20*, 367–387.

(19) Feynman, R. P. *Science* **1966**, *153*, 699–708.

(20) Kac, M. *Probability and related topics in physical sciences*; Interscience Publishers: London, New York, 1959; Chapter IV, p xiii, 266 p.

(21) Kac, M. *Trans. Am. Math. Soc.* **1949**, *65*, 1–13.

(22) Chaichian, M.; Demichev, A. P. *Path integrals in physics*. Philadelphia, PA, Bristol, U.K., 2001.

(23) Schulman, L. S. *Techniques and applications of path integration*; Wiley: New York, 1981; p xv, 359 p.

(24) Dirac, P. A. M. *The principles of quantum mechanics*, 4th ed.; Clarendon Press: Oxford, England, 1981; p xii, 314 p. For the action-principle in quantum mechanics, see Section 32, p 125.

(25) (a) Dirac, P. A. M. Phys. Z. Sowjetunion 1933, Band*3*, 64−72. (b) English translation: Brown, L. M. *Feynman's thesis: a new approach to quantum theory*; World Scientific: Singapore, Hackensack, NJ, 2005;. pp 111−119.

(26) Gutzwiller, M. C *Am. J. Phys* **1998**, *66*, 304–324. For path integrals, see Sections VII. C and VIII. C.

(27) Mielke, S. L.; Truhlar, D. G. *J. Chem. Phys.* **2001**, *115*, 652–662.

(28) McQuarrie, D. A. *Statistical mechanics*; University Science Books: Sausalito, CA, 2000; p xii, 641 p.

(29) (a) Sauer, T. In *Fluctuating paths and fields: festschrift dedicated to Hagen Kleinert on the occasion of his 60th birthday*; Janke, W., Pelster, A., Schmidt, H.-J., Bachmann, M., Eds.; World Scientific: River Edge, NJ, 2001; pp 29−42. (b) Los Alamos National Laboratory, Preprint Archive, Physics arXiv:physics/0107010v1 [physics.hist-ph], 1 (2001).

(30) Fosdick, L. D. *J. Math. Phys.* **1962**, *3*, 1251–1264.

(31) Fosdick, L. D. *SIAM Rev.* **1968**, *10*, 315–328.

(32) Morita, T. *J. Phys. Soc. Jpn.* **1973**, *35*, 980–4.

(33) Barker, J. A. *J. Chem. Phys.* **1979**, *70*, 2914–18.

(34) MacKeown, P. K. *Am. J. Phys* **1985**, *53*, 880–885.

(35) Chandler, D.; Wolynes, P. G. *J. Chem. Phys.* **1981**, *74*, 4078–95.

(36) Berne, B. J.; Thirumalai, D. *Annu. Rev. Phys. Chem.* **1986**, *37*, 401–424.

(37) Ceperley, D. M. *Rev. Mod. Phys.* **1995**, *67*, 279–355.

(38) Mielke, S. L.; Truhlar, D. G. *J. Chem. Phys.* **2001**, *114*, 621–630.

(39) Coalson, R. D. *J. Chem. Phys.* **1986**, *85*, 926–936.

(40) Cao, J.; Voth, G. A. *J. Chem. Phys.* **1994**, *100*, 5093–105.

(41) Cao, J.; Voth, G. A. *J. Chem. Phys.* **1994**, *100*, 5106–18.

(42) Cao, J.; Voth, G. A. *J. Chem. Phys.* **1994**, *101*, 6157–67.

(43) Cao, J.; Voth, G. A. *J. Chem. Phys.* **1994**, *101*, 6184–92.

(44) Marx, D.; Parrinello, M. *Nature (London)* **1995**, *375*, 216–18.

(45) Tuckerman, M. E.; Marx, D.; Klein, M. L.; Parrinello, M. *Science* **1997**, *275*, 817–820.

(46) Tuckerman, M. E.; Marx, D.; Parrinello, M. *Nature (London)* **2002**, *417*, 925–929.

(47) Marx, D.; Tuckerman, M. E.; Martyna, G. J. *Comput. Phys. Commun.* **1999**, *118*, 166–184.

(48) Paesani, F.; Iuchi, S.; Voth, G. A. *J. Chem. Phys.* **2007**, *127*, 074506.

(49) Ohta, Y.; Ohta, K.; Kinugawa, K. *J. Chem. Phys.* **2004**, *120*, 312–320.

(50) Hayashi, A.; Shiga, M.; Tachikawa, M. *J. Chem. Phys.* **2006**, *125*, 204310.

(51) Gao, J.; Wong, K.-Y.; Major, D. T. *J. Comput. Chem.* **2008**, *29*, 514–522.

(52) Major, D. T.; Gao, J. *J. Am. Chem. Soc.* **2006**, *128*, 16345–16357.

(53) Gao, J.; Major, D. T.; Fan, Y.; Lin, Y.-l.; Ma, S.; Wong, K.-Y. In *Molecular Modeling of Proteins*; Kukol, A., Ed.; Springer Verlag: 2008; pp 37−62.

(54) Chakrabarti, N.; Carrington, T., Jr.; Roux, B. *Chem. Phys. Lett.* **1998**, *293*, 209–220.

(55) Field, M. J.; Albe, M.; Bret, C.; Martin, F. P.-D.; Thomas, A. *J. Comput. Chem.* **2000**, *21*, 1088–1100.

(56) Warshel, A.; Olsson, M. H. M.; Villa-Freixa, J. In *Isotope Effects in Chemistry and Biology*; Kohen, A., Limbach, H.-H., Eds.; Taylor & Francis: Boca Raton, 2006; pp 621−644.

(57) Wang, M.; Lu, Z.; Yang, W. *J. Chem. Phys.* **2006**, *124*, 124516.

(58) Wang, Q.; Hammes-Schiffer, S. *J. Chem. Phys.* **2006**, *125*, 184102.

(59) Wong, K.-Y.; Gao, J. *J. Chem. Phys.* **2007**, *127*, 211103.

(60) Kleinert, H. *Phys. Rev. D* **1998**, *57*, 2264–2278.

(61) Janke, W.; Kleinert, H. *Phys. Rev. Lett.* **1995**, *75*, 2787–91.

(62) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab initio molecular orbital theory*; Wiley: New York, 1986; p xviii, 548 p.

(63) Szabo, A.; Ostlund, N. S. *Modern quantum chemistry: introduction to advanced electronic structure theory*; 2nd ed.; Dover Publications: Mineola, NY, 1996; p xiv, 466 p. For the Rayleigh−Schrödinger perturbation theory, see section 6.1, p 322.

(64) Helgaker, T.; Jørgensen, P.; Olsen, J. *Molecular electronic-structure theory*; Wiley: Chichester, NY, 2000; p xxvii, 908 p.

(65) Parr, R. G.; Yang, W. *Density-functional theory of atoms and molecules*; Oxford University Press: Clarendon Press: New York, NY, Oxford, U.K., 1989; p x, 333 p.

(66) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

(67) Giachetti, R.; Tognetti, V. *Phys. Rev. Lett.* **1985**, *55*, 912–15.

(68) Feynman, R. P.; Kleinert, H. *Phys. Rev. A* **1986**, *34*, 5080–5084.

(69) Voth, G. A.; Chandler, D.; Miller, W. H. *J. Chem. Phys.* **1989**, *91*, 7749–60.

(70) Weissbach, F.; Pelster, A.; Hamprecht, B. *Phys. Rev. E* **2002**, *66*, 036129.

(71) Byrnes, T. M. R.; Hamer, C. J.; Zheng, W.; Morrison, S. *Phys. Rev. D* **2003**, *68*, 016002.

(72) Filinov, A. V.; Golubnychiy, V. O.; Bonitz, M.; Ebeling, W.; Dufty, J. W. *Phys. Rev. E* **2004**, *70*, 046411.

(73) Palmieri, B.; Ronis, D. *Phys. Rev. E* **2006**, *73*, 061105.

(74) Srivastava, S.; Vishwamittar *Phys. Rev. A* **1991***44*, 8006−8011.

(75) Sesé, L. M. *Mol. Phys.* **1999**, *97*, 881–896.

(76) Poulsen, J. A.; Nyman, G.; Rossky, P. J. *J. Chem. Phys.* **2003**, *119*, 12179–12193.

(77) Poulsen, J. A.; Nyman, G.; Rossky, P. J. *J. Chem. Theory Comput.* **2006**, *2*, 1482–1491.

(78) Poulsen, J. A.; Scheers, J.; Nyman, G.; Rossky, P. J. *Phys. Rev. B* **2007**, *75*, 224505.

(79) Coker, D. F.; Bonella, S. In *Quantum Dynamics of Complex Molecular Systems, Springer Series in Chemical Physics*; Micha, D. A., Burghardt, I., Eds.; Springer: New York, 2007; Vol. 83, pp 321−342.

(80) Morse, P. M. *Phys. Rev.* **1929**, *34*, 57–64.

(81) Eckart, C. *Phys. Rev.* **1930**, *35*, 1303–1309.

(82) MacDonald, J. K. L. *Phys. Rev.* **1933**, *43*, 830–833.

(83) Rayleigh, J. W. S. *Philos. Trans. R. Soc. London* **1871**, *161*, 77.

(84) Ritz, W. *J. Reine Angew. Math.* **1908**, *135*, 1–61.

(85) Arfken, G. B.; Weber, H.-J. *Mathematical methods for physicists*, 5th ed.; Academic Press: San Diego, 2001; p xiv, 1112 p. Section 17.8, p 1052.

(86) Rayleigh, J. W. S. *The theory of sound*, 2nd ed.; Dover: New York, 1945; Vol. 1−2, p Vol. 1: xlii, 480 p; Vol. 2: xvi, 504 p.

(87) (a) Schrödinger, E. *Ann. Phys.* **1926**, *80*, 437–476. (b) English translation: Schrödinger, E. *Collected papers on wave mechanics: together with his Four lectures on wave mechanics*, 3rd ed.; Chelsea Publishing: New York, 1982.

(88) Bell, R. P. *The tunnel effect in chemistry*; Chapman and Hall: London, New York, 1980; p ix, 222 p.

(89) (a) Cha, Y.; Murray, C. J.; Klinman, J. P. *Science* **1989**, *243*, 1325–30. (b) Erratum : Cha, Y.; Murray, C. J.; Klinman, J. P. *Science* **1989**, *244*, 1030.

(90) Tanaka, H.; Kanoh, H.; Yudasaka, M.; Iijima, S.; Kaneko, K. *J. Am. Chem. Soc.* **2005**, *127*, 7511–7516.

(91) Doll, J. D.; Myers, L. E. *J. Chem. Phys.* **1979**, *71*, 2880–3.

(92) Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22*, 1420–1426.

(93) Jorgensen, W. L. *Acc. Chem. Res.* **1989**, *22*, 184–9.

(94) Kollman, P. *Chem. Rev.* **1993**, *93*, 2395–417.

(95) Kubo, R. *J. Phys. Soc. Jpn.* **1962**, *17*, 1100–1120.

(96) Stratt, R. M. *Acc. Chem. Res.* **1995**, *28*, 201–7.

(97) Deng, Y.; Ladanyi, B. M.; Stratt, R. M. *J. Chem. Phys.* **2002**, *117*, 10752–10767.

(98) Wolfram Research, Inc., Mathematica, Versions 5 and 6, Champaign, IL.

(99) Flügge, S. *Practical quantum mechanics*; Springer Verlag: New York, 1974; p xv, 331, 287 p. Problem 70, p 182.

(100) Major, D. T.; Gao, J. *J. Mol. Graphics Modell.* **2005**, *24*, 121–127.

(101) Jang, S.; Schwieters, C. D.; Voth, G. A. *J. Phys. Chem. A* **1999**, *103*, 9527–9538.

(102) Jang, S.; Voth, G. A. *J. Chem. Phys.* **2000**, *112*, 8747–8757.

(103) Johnston, H. S. *Gas phase reaction rate theory*; Ronald Press Co.: New York, 1966; p ix, 362 p. For the Eckart potential, see Chapter 2. For the corrections to the errata in Chapter 2, see Notes [1] in Garrett, B. C.; Truhlar, D. G. *J. Phys. Chem.* **1979**, *83*, 2921–2926.

(104) Kreevoy, M. M.; Truhlar, D. G. In *Techniques of chemistry: Investigation of rates and mechanisms of reactions*, 4th ed.; Bernasconi, C. F., Ed.; Wiley: New York, 1986; Vol. 6, pp 13−95.

(105) Kohen, A.; Limbach, H.-H. *Isotope effects in chemistry and biology*; Taylor & Francis: Boca Raton, 2006; p xiv, 1074 p.

CT800109S

JCTC Journal of Chemical Theory and Computation

Stability and Dissociation Energies of Open-Chain N_4C_2

Kasha Casey, Jessica Thomas, Kiara Fairman, and Douglas L. Strout*

Department of Physical Sciences, Alabama State University,
Montgomery, Alabama 36101

Received April 2, 2008

Abstract: Complex forms of nitrogen are of interest due to their potential as high-energy materials. Many forms of nitrogen, including open-chain and cage molecules, have been studied previously. While many all-nitrogen molecules N_x have been shown to be too unstable for high-energy applications, it has been shown that certain heteroatoms (including carbon) can stabilize a nitrogen structure. A molecule that is not 100% nitrogen will be less energetic, but that energy loss is a tradeoff for the improved stability. In this study, open-chain N_4C_2 (70% nitrogen by mass) isomers are studied by theoretical calculations to determine isomer stability and dissociation energies. Calculations are carried out with density functional theory (PBE1PBE), perturbation theory (MP2), and coupled-cluster theory (CCSD(T)). Trends in stability of the molecules are calculated and discussed.

Introduction

Nitrogen molecules have been the subjects of many recent studies because of their potential as high energy density materials (HEDM). An all-nitrogen molecule N_x can undergo the reaction $N_x \rightarrow (x/2)N_2$, a reaction that can be exothermic by 50 kcal/mol or more per nitrogen atom.[1,2] To be a practical energy source, however, a molecule N_x would have to resist dissociation well enough to be a stable fuel. Theoretical studies[3–7] have shown that numerous N_x molecules are not sufficiently stable to be practical HEDM, including cyclic and acyclic isomers with eight to twelve atoms. Cage isomers of N_8 and N_{12} have also been shown[7–10] by theoretical calculations to be unstable. Experimental progress in the synthesis of nitrogen molecules has been very encouraging, with the N_5^+ and N_5^- ions having been recently produced[11,12] in the laboratory. More recently, a network polymer of nitrogen has been produced[13] under very high pressure conditions. Experimental successes have sparked theoretical studies[1,14,15] on other potential all-nitrogen molecules. More recent developments include the experimental synthesis of high energy molecules consisting predominantly of nitrogen, including azides[16,17] of various molecules and polyazides[18,19] of atoms and molecules, such as 1,3,5-triazine. Future developments in experiment and theory will further broaden the horizons of high energy nitrogen research.

The stability properties of N_x molecules have also been extensively studied in a computational survey[20] of various structural forms with up to 20 atoms. Cyclic, acyclic, and cage isomers have been examined to determine the bonding properties and energetics over a wide range of molecules. A more recent computational study[21] of cage isomers of N_{12} examined the specific structural features that lead to the most stable molecules among the three-coordinate nitrogen cages. Those results showed that molecules with the most pentagons in the nitrogen network tend to be the most stable, with a secondary stabilizing effect due to triangles in the cage structure. A recent study[22] of larger nitrogen molecules N_{24}, N_{30}, and N_{36} showed significant deviations from the pentagon-favoring trend. Each of these molecule sizes has fullerene-like cages consisting solely of pentagons and hexagons, but a large stability advantage was found for molecules with fewer pentagons, more triangles, and an overall structure more cylindrical than spheroidal. Studies[23,24] of intermediate-sized molecules N_{14}, N_{16}, and N_{18} also showed that the cage isomer with the most pentagons was not the most stable cage, even when compared to isomer(s) containing triangles (which have 60° angles that should have significant angle strain). For each of these molecule sizes, spheroidally shaped molecules proved to be less stable than elongated, cylindrical ones.

However, while it is possible to identify in relative terms which nitrogen cages are the most stable, it has been shown[7]

* Corresponding author. E-mail: dstrout@alasu.edu.

10.1021/ct8001943 CCC: $40.75 © 2008 American Chemical Society
Published on Web 07/29/2008

in the case of N_{12} that even the most stable N_{12} cage is unstable with respect to dissociation. The number of studies demonstrating the instability of various all-nitrogen molecules has resulted in considerable attention toward compounds that are predominantly nitrogen but contain heteroatoms that stabilize the structure. In addition to the experimental studies[16–18] cited above, theoretical studies have been carried out that show, for example, that nitrogen cages can be stabilized by oxygen insertion[25,26] or phosphorus substitution.[27]

A study[28] of carbon–nitrogen cages showed that carbon substitution into an N_{12} cage results in a stable $N_6C_6H_6$, but the only isomer considered was one in which the six carbon atoms replaced the nitrogen atoms in the two axial triangles of the original N_{12}. A further study[29] of several isomers of $N_6C_6H_6$ showed that, for substitutions of C–H bonding groups into an N_{12} cage, the most stable isomers were the ones with the largest number of C–N bonds. Also, the isomers with the highest number of C–N bonds also had the highest dissociation energies in the N–N bonds, which is significant because the N–N were weaker than other bonds in the cage. The strength of the N–N bonds, therefore, plays a key role in the overall stability of the molecules with respect to dissociation. Similar studies[30] have been carried out for cage isomers of $N_8C_8H_8$.

In the current study, a similar idea is applied to open-chain molecules rather than cages. Isomers of open-chain N_4C_2 are studied as two carbon atoms are substituted into the structure of open-chain N_6. Variations in the placement of the two carbon atoms are considered and discussed. For nomenclature purposes, each isomer shall be named according to the positions of the two carbons. "Isomer 12", for example, shall refer to a chain of six atoms with positions 1 and 2 occupied by carbon and the rest nitrogen atoms. The energies of these isomers are calculated with respect to each other and with respect to plausible dissociation products, thereby giving a measure of stability with respect to dissociation.Figure 1.

Computational Methods

Geometries are optimized with second-order Moller–Plesset perturbation theory[31] (MP2) and density functional theory[32] (PBE1PBE). Single energy points are calculated with coupled-cluster theory[33] (CCSD(T)). Each molecule in this study is calculated in its own ground state, which is the singlet for N_4C_2 but not necessarily the singlet for all dissociation products. The correlation-consistent basis sets[34] of Dunning are used, including double-ζ (cc-pVDZ), augmented double-ζ (aug-cc-pVDZ), and triple-ζ (cc-pVTZ). The Gaussian 03 computational chemistry software[35] and Windows counterpart Gaussian 03W are used for the calculations in this study.

Results and Discussion

Isomer stability trends. Open-chain N_4C_2 has nine structural isomers, designated by the carbon positions as isomers 12, 13, 14, 15, 16, 23, 24, 25, and 34, respectively. PBE1PBE, MP2, and CCSD(T) relative energies with the cc-pVDZ for the nine isomers are shown in Table 1. Two trends in stability appear in the data. The primary trend appears to be a disfavoring of isomers with a carbon atom on the end of the chain. Most of the isomers with a carbon on the end are at the high end of the energy range, and even isomer 15 is much less stable than the structurally similar isomer 25. Mulliken charges shown in Table 1 show a correlation between stability and the charges on the carbon atoms. Carbon bonded to nitrogen will typically take a positive partial charge because of differences in electronegativity, but carbon on the end of the N_4C_2 chain tends to take on a partial negative charge. (If the atom on the end of the chain has a triple bond and a lone pair, this leads to neutral nitrogen but negative carbon.) The positive charges on carbon in isomer 25, and their interactions with neighboring negative nitrogen, stabilize isomer 25 and cause it to be the most stable N_4C_2 chain.

The secondary trend is a disfavoring of isomers with neighboring carbon atoms. Isomers with neighboring carbon atoms have one carbon–carbon bond, two carbon–nitrogen bonds, and two nitrogen–nitrogen bonds. Isomers without neighboring carbon atoms have zero carbon–carbon bonds, four carbon–nitrogen bonds, and one nitrogen–nitrogen

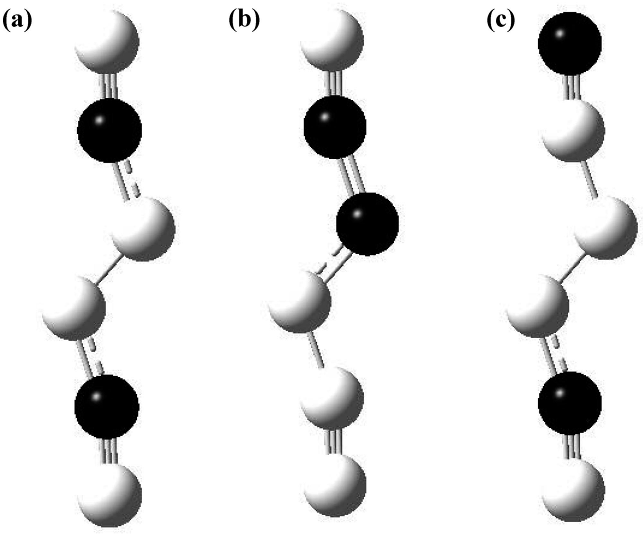


Figure 1. (a) N_4C_2 isomer 25 (C_{2h} point group symmetry). (b) N_4C_2 isomer 23 (C_s point group symmetry). (c) N_4C_2 isomer 15 (C_s point group symmetry). Carbon atoms are shown in black; nitrogen atoms are in white.

Table 1. PBE1PBE, MP2, and CCSD(T)/MP2 Relative Energies with the cc-pVDZ Basis Set for Isomers of Open-Chain N_4C_2 (Energies in kilocalories/mole)[a]

isomer	PBE1PBE	MP2	CCSD(T)/MP2	carbon charges
25	0.0	0.0	0.0	+0.19, +0.19
23	+28.0	+31.9	+33.2	+0.02, −0.03
15	+29.3	+34.9	+30.2	−0.03, +0.18
34	+34.4	+35.9	+42.6	−0.31, −0.31
13	+42.0	+49.8	+47.7	−0.09, +0.02
14	+43.4	*b*		−0.11, +0.02
16	+61.4	+71.8	+63.1	−0.04, −0.04
12	+92.0	+106.3	+100.4	−0.15, +0.08
24	*b*			

[a] Mulliken charges on the carbon atoms are included. [b] Optimization was dissociative (loss of N_2).

Table 2. Cis-Trans Energies for N_4C_2 Isomers 25, 15, and 16 Calculated with the PBE1PBE/cc-pVDZ Method (Relative Energies in kilocalories/mole)

isomer	trans	cis
25	0.0	+4.3
15	0.0	+2.6
16	0.0	+0.6

bond. Therefore, the introduction of neighboring carbon results in the replacement of two carbon−nitrogen bonds with carbon−carbon and nitrogen−nitrogen bonds. That replacement results in a net reduction in bond enthalpy and a net loss in overall stability. The least stable isomer in this study, isomer 12, has both destabilizing features; one carbon is on the end of the chain, and the two carbon atoms are neighbors.

Cis−Trans Isomers. From a chemical point of view, isomers 25, 15, and 16 can be viewed as azo compounds because of the central N—N bonding group, but do these compounds really have the structure of azo compounds? The Lewis diagram of isomer 25, for example, would show a central N═N double bond, C—N single bonds to the azo bonding group, and chain terminal C—N triple bonds in the cyano groups. Modeling these bonds by, respectively, *trans*-N_2H_2, methylamine, and HCN yields bond lengths of 1.239 Å for N═N, 1.454 Å for C—N single bond, and 1.156 Å for C—N triple bond, at the PBE1PBE/cc-pVDZ level of theory. The corresponding bond lengths for isomer 25 are 1.259 Å for N═N (slightly elongated), 1.355 Å for C—N single bond (significantly shortened), and 1.165 Å for C≡N triple bond (slightly elongated). Since the shortening of the C—N single bonds is likely the result of resonance delocalization from conjugated π-bonds, the structure of isomer 25 is very consistent with a trans-azo compound with cyano bonding groups.

This opens the possibility of cis−trans isomerism with respect to the cyano and isocyano bonding groups attached to the azo center. All of the energies in Table 1 involve trans isomers of 25, 15, and 16. However, to verify that these trans isomers are the most stable form, cis isomers have been optimized with the PBE1PBE/cc-pVDZ method, and the results are shown in Table 2. The calculations confirm that the trans isomers are, in fact, more stable than their cis counterparts, by an amount of energy that follows the stability of the isomers themselves. The most stable isomer, 25, has a cis−trans gap of 4.3 kcal/mol, whereas the highly unstable isomer 16 has a cis−trans gap of less than 1 kcal/mol.

What happened to isomer 24? The arguments in favor of the stability of isomer 25, namely, the lack of chain-terminal carbon and the lack of carbon−carbon bonds, should apply equally well to isomer 24. Yet, no data exists for isomer 24 by reason that its PBE1PBE/cc-pVDZ geometry optimization failed. Since the optimization problem was the uncontrolled lengthening of the bond between atoms 4 and 5, detaching N_2 from the end of the chain, a more detailed study of this bond was conducted. A series of PBE1PBE/cc-pVDZ optimizations was carried out with frozen values for the 4−5 bond ranging from 1.30 to 1.55 Å. The optimized energies from these calculations are comparable to the most stable isomers in this study, but in all cases, the first derivative of the energy with respect to the 4−5 bond indicated an energetic preference for lengthening. Therefore, in a totally unconstrained optimization, this bond will lengthen indefinitely without finding a bound local minimum.

Table 3. Dissociation Energies of N_4C_2 Isomer 25 (Energies in kilocalories/mole)

		products		
		NCN_2 + CN	NCN + NCN	CN + CN + N_2
PBE1PBE	cc-pVDZ	+85.8	+61.0	+87.6
MP2	cc-pVDZ	+110.9	+113.8	+100.4
CCSD(T)/MP2	cc-pVDZ	+83.4	+63.3	+65.7
MP2	aug-cc-pVDZ	+105.6	+117.7	+102.9
CCSD(T)/MP2	aug-cc-pVDZ	+77.5	+65.4	+68.5
MP2	cc-pVTZ	+107.2	+121.0	+104.8
CCSD(T)/MP2	cc-pVTZ	+81.0	+69.9	+71.8

Table 4. Dissociation Energies of N_4C_2 Isomer 23 (Energies in kilocalories/mole)

		Products		
		NC_2N + N_2	NC_2 + N_3	CN + CN + N_2
PBE1PBE	cc-pVDZ	−87.8	+92.4	+59.6
MP2	cc-pVDZ	−111.8	+129.5	+68.5
CCSD(T)/MP2	cc-pVDZ	−108.0	+93.7	+32.5
MP2	aug-cc-pVDZ	−109.2	+133.1	+71.0
CCSD(T)/MP2	aug-cc-pVDZ	−105.2	+96.1	+35.7
MP2	cc-pVTZ	−109.6	+121.0	+73.2
CCSD(T)/MP2	cc-pVTZ	−106.4	+97.3	+39.3

Table 5. Dissociation energies of N_4C_2 isomer 15 (Energies in kilocalories/mole)

		products		
		NCN_2 + CN	NCN + CN_2	CN + CN + N_2
PBE1PBE	cc-pVDZ	+56.5	+59.4	+58.3
MP2	cc-pVDZ	+76.0	+101.5	+65.5
CCSD(T)/MP2	cc-pVDZ	+53.2	+62.8	+35.5
MP2	aug-cc-pVDZ	+70.7	+105.6	+67.9
CCSD(T)/MP2	aug-cc-pVDZ	+47.3	+64.9	+38.3
MP2	cc-pVTZ	+72.6	+109.0	+70.2
CCSD(T)/MP2	cc-pVTZ	+50.5	+69.0	+41.4

Dissociation Energies. The three most thermodynamically stable isomers, namely 25, 15, and 23, are subjected to dissociation studies, and the results are shown in Tables 3, 4 and 5 for isomers 25, 23, and 15, respectively. (Both binary and trinary dissociations are included, because of completeness and because the CN_3 dissociation product itself dissociated to CN and N_2 anyway upon optimization.) CCSD(T)/MP2 predicts lower dissociation energies than MP2, and in most cases, basis set effects cause small increases in dissociation energy. Generally speaking, the PBE1PBE dissociation energies are more accurate than MP2, as compared to CCSD(T). Isomer 23 has a very exothermic dissociation ($N_4C_2 \rightarrow NC_2N + N_2$) whereby the molecule loses N_2 from the end of the chain. This behavior is similar to previously studied behavior of all-nitrogen chains. All-nitrogen chains lose N_2 very easily, which is why they are kinetically unstable. Isomer 23 is similarly unstable with

Table 6. PBE1PBE/cc-pVDZ Detonation Energies for N_4C_2 Isomers 25, 23, and 15

isomer	energy (kcal/mol)	energy (kcal/g)
25	132.9	1.66
23	161.0	2.01
15	162.3	2.03

respect to dissociation and is therefore not likely to hold promise as a high-energy material.

Isomers 25 and 15 have dissociation processes that are all endothermic, meaning that these molecules have more resistance to dissociation than isomer 23. For isomer 15, the lowest energy dissociation involves the loss of one or both CN units from the end of the molecule. CCSD(T)/MP2 predicts that dissociation of isomer 15 costs at least 40−50 kcal/mol of energy, and therefore, isomer 15 may have potential as high-energy material. The stability of isomer 25 is even greater than that of isomer 15, with all dissociations requiring at least 70 kcal/mol at the CCSD(T)/cc-pVTZ level of theory. Because of the structural similarities between isomers 25 and 15, most of the dissociation products are the same for the two molecules. The greater kinetic stability of isomer 25 versus isomer 15, reflected in Tables 2 and 3, is a direct result of the greater thermodynamic stability shown in Table 1.

Detonation Energies. Detonation energies of the molecules are calculated[36] as the energy released by the reaction $N_4C_2 \rightarrow 2N_2 + 2C$ (graphite). Since N_2 and graphitic carbon are the most stable allotropes of the products, these detonation energies also indicate the heats of formation of the N_4C_2 chains. The energies for isomers 25, 23, and 15 are shown in Table 6. Since all these reactions lead to the same products, the differences in energy are the same as the differences in stability shown in Table 1. A CCSD(T)/cc-pVDZ estimate, based on the model reaction $2HCN + N_2H_4 \rightarrow 3H_2 + N_4C_2$, of the heat of formation[37] of isomer 25 predicts a value of 154 kcal/mol, as compared with the 133 kcal/mol in Table 6, possibly indicating that the values in Table 6 are underestimating the detonation energy.

Since isomer 25 has the lowest energy (most stable) to start with, it has the lowest energy of detonation. According to Table 6, isomer 25 has the ability to release about 1.7 kcal/g, while isomers 23 and 15 have an energy release of about 2 kcal/g. The energy release occurs principally because of the formation of N_2, so the fact that the molecules are 70% nitrogen by mass is a favorable energetic property. These detonation energies are comparable to, for example, the previously studied $N_6C_6H_6$ cages.[29] The difference between these open chains and previously studied cages is that the cages have single bonds, whereas the open chains have double and triple bonds, which are less energetic upon decomposition to N_2. That is why open chains that are 70% nitrogen have about the same detonation energy as cages that are slightly more than 50% nitrogen. If stable open chains can be found that are even richer in nitrogen, the energetic properties will improve accordingly.

Conclusion

N_4C_2 is an example of a carbon-substituted nitrogen chain that could hold promise as a high-energy material, but there are important considerations in the design of such a molecule for optimal high-energy properties. Carbon atoms should be distributed along a chain such that no C−C bonds occur, but the carbon atoms should not occupy chain-terminal positions. On the other hand, chain-terminal N_2 is likely to be easily lost from a chain, so the best arrangement of carbon atoms would include carbon atoms at the second position with respect to each end of the chain (as seen in isomer 25). N_4C_2 is 70% nitrogen by mass and has favorable energetic properties. Longer chains that are richer in nitrogen will be even better energetic materials, if they are stable.

Acknowledgment. The Alabama Supercomputer Authority is gratefully acknowledged for a grant of computer time on the SGI Altix in Huntsville, AL. This work was partially supported by the National Computational Science Alliance under grant number CHE050022N and utilized the IBM p690 cluster in Champaign, Illinois. This work was supported by the National Science Foundation (NSF/HBCU-UP grant 0505872). K.C., J.T., and K.F. are supported by NSF/HBCU-UP as undergraduate scholars. This work was also supported by the National Institutes of Health (NIH/NCMHD 1P20MD000547-01) and the Petroleum Research Fund, administered by the American Chemical Society (PRF 43798-B6). The taxpayers of the state of Alabama in particular and the United States in general are gratefully acknowledged.

Supporting Information Available: PBE1PBE/cc-pVDZ energies and geometries, as well as structural parameters and singlet−triplet excitation energies. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Fau, S.; Bartlett, R. J. *J. Phys. Chem. A* **2001**, *105*, 4096.

(2) Tian, A.; Ding, F.; Zhang, L.; Xie, Y.; Schaefer, H. F. *J. Phys. Chem. A* **1997**, *101*, 1946.

(3) Chung, G.; Schmidt, M. W.; Gordon, M. S. *J. Phys. Chem. A* **2000**, *104*, 5647.

(4) Strout, D. L. *J. Phys. Chem. A* **2002**, *106*, 816.

(5) Thompson, M. D.; Bledson, T. M.; Strout, D. L. *J. Phys. Chem. A* **2002**, *106*, 6880.

(6) (a) Li, Q. S.; Liu, Y. D. *Chem. Phys. Lett.* **2002**, *353*, 204. (b) Li, Q. S.; Qu, H.; Zhu, H. S. *Chin. Sci. Bull.* **1996**, *41*, 1184.

(7) (a) Li, Q. S.; Zhao, J. F. *J. Phys. Chem. A* **2002**, *106*, 5367. (b) Qu, H.; Li, Q. S.; Zhu, H. S. *Chin. Sci. Bull.* **1997**, *42*, 462.

(8) Gagliardi, L.; Evangelisti, S.; Widmark, P. O.; Roos, B. O. *Theor. Chem. Acc.* **1997**, *97*, 136.

(9) Gagliardi, L.; Evangelisti, S.; Bernhardsson, A.; Lindh, R.; Roos, B. O. *Int. J. Quantum Chem.* **2000**, *77*, 311.

(10) Schmidt, M. W.; Gordon, M. S.; Boatz, J. A. *Int. J. Quantum Chem.* **2000**, *76*, 434.

(11) Christe, K. O.; Wilson, W. W.; Sheehy, J. A.; Boatz, J. A. *Angew. Chem., Int. Ed.* **1999**, *38*, 2004.

(12) (a) Vij, A.; Pavlovich, J. G.; Wilson, W. W.; Vij, V.; Christe, K. O. *Angew. Chem., Int. Ed.* **2002**, *41*, 3051. (b) Butler, R. N.; Stephens, J. C.; Burke, L. A. *Chem. Commun.* **2003**, *8*, 1016.

(13) Eremets, M. I.; Gavriliuk, A. G.; Trojan, I. A.; Dzivenko, D. A.; Boehler, R. *Natur. Mater.* **2004**, *3*, 558.

(14) Fau, S.; Wilson, K. J.; Bartlett, R. J. *J. Phys. Chem. A* **2002**, *106*, 4639.

(15) Dixon, D. A.; Feller, D.; Christe, K. O.; Wilson, W. W.; Vij, A.; Vij, V.; Jenkins, H. D. B.; Olson, R. M.; Gordon, M. S. *J. Am. Chem. Soc.* **2004**, *126*, 834.

(16) Knapp, C.; Passmore, J. *Angew. Chem., Int. Ed.* **2004**, *43*, 4834.

(17) Haiges, R.; Schneider, S.; Schroer, T.; Christe, K. O. *Angew. Chem., Int. Ed.* **2004**, *43*, 4919.

(18) Huynh, M. V.; Hiskey, M. A.; Hartline, E. L.; Montoya, D. P.; Gilardi, R. *Angew. Chem., Int. Ed.* **2004**, *43*, 4924.

(19) (a) Klapotke, T. M.; Schulz, A.; McNamara, J. *J. Chem. Soc., Dalton Trans.* **1996**, 2985. (b) Klapotke, T. M.; Noth, H.; Schutt, T.; Warchhold, M. *Angew. Chem., Int. Ed.* **2000**, *39*, 2108. (c) Klapotke, T. M.; Krumm, R.; Mayer, P.; Schwab, I. *Angew. Chem., Int. Ed.* **2003**, *42*, 5843.

(20) Glukhovtsev, M. N.; Jiao, H.; Schleyer, P. v. R. *Inorg. Chem.* **1996**, *35*, 7124.

(21) Bruney, L. Y.; Bledson, T. M.; Strout, D. L. *Inorg. Chem.* **2003**, *42*, 8117.

(22) Strout, D. L. *J. Phys. Chem. A* **2004**, *108*, 2555.

(23) Sturdivant, S. E.; Nelson, F. A.; Strout, D. L. *J. Phys. Chem. A* **2004**, *108*, 7087.

(24) Strout, D. L. *J. Phys. Chem. A* **2004**, *108*, 10911.

(25) Strout, D. L. *J. Phys. Chem. A* **2003**, *107*, 1647.

(26) Sturdivant, S. E.; Strout, D. L. *J. Phys. Chem. A* **2004**, *108*, 4773.

(27) Strout, D. L. *J. Chem. Theory Comput.* **2005**, *1*, 561.

(28) Colvin, K. D.; Cottrell, R.; Strout, D. L. *J. Chem. Theory Comput.* **2006**, *2*, 25.

(29) Strout, D. L. *J. Phys. Chem. A* **2006**, *110*, 7228.

(30) Cottrell, R.; McAdory, D.; Jones, J.; Gilchrist, A.; Shields, D.; Strout, D. L. *J. Phys. Chem. A* **2006**, *110*, 13889.

(31) Moller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618.

(32) (a) Perdew, J. P.; Ernzerhof, M. *J. Chem. Phys.* **1996**, *105*, 9982. (b) Ernzerhof, M.; Scuseria, G. E. *J. Chem. Phys.* **1999**, *110*, 5029. (c) Adamo, C.; Barone, V. *J. Chem. Phys.* **1999**, *110*, 6158.

(33) (a) Purvis, G. D.; Bartlett, R. J. *J. Chem. Phys.* **1982**, *76*, 1910. (b) Scuseria, G. E.; Janssen, C. L.; Schaefer, H. F. *J. Chem. Phys.* **1988**, *89*, 7382.

(34) Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007.

(35) 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.

(36) The per-atom energy of graphitic carbon was calculated using graphitic model molecules $C_{24}H_{12}$ and $C_{54}H_{18}$. If E(coronene) $= 12E_C + 12E_{CH}$ and $E(C_{54}H_{18}) = 36E_C + 18E_{CH}$, where E_C and E_{CH} represent interior and edge carbon, respectively, then E_C is the per-atom energy of graphitic carbon.

(37) Heats of formation for HCN(g) and N_2H_4(g) are taken from Dean, J. A. *Lange's Handbook of Chemistry*, 14th ed.; McGraw-Hill: New York, 1992.

CT8001943

JCTC Journal of Chemical Theory and Computation

Dissecting the Hydrogen Bond: A Quantum Monte Carlo Approach

Fabio Sterpone,[†,⊥] Leonardo Spanu,[‡,⊥,#] Luca Ferraro,[†] Sandro Sorella, and Leonardo Guidoni*[,§,||]

CASPUR, Via dei Tizii 6B, 00185, Roma, Italy, International School for Advanced Studied (SISSA/ISAS), Via Beirut 4, 34014 Trieste, Italy, Dipartimento di Fisica, La Sapienza - Universita di Roma, P.le A. Moro 2, 00185 Roma, Italy, and NAST Centre - Nanoscience & Nanotechnology & Instrumentation, Università degli Studi di Roma Tor Vergata, Roma, Italy

Received April 7, 2008

Abstract: We present a Quantum Monte Carlo study of the dissociation energy and the dispersion curve of the water dimer, a prototype of hydrogen bonded system. Our calculations are based on a wave function which is a modern and fully correlated implementation of the Pauling's valence bond idea: the Jastrow Antisymmetrised Geminal Power (JAGP) [Casula et al. *J. Chem. Phys.* **2003**, *119*, 6500−6511]. With this variational wave function we obtain a binding energy of −4.5(0.1) kcal/mol that is only slightly increased to −4.9(0.1) kcal/mol by using the Lattice Regularized Diffusion Monte Carlo (LRDMC). This projection technique allows for the substantial improvement in the correlation energy of a given variational guess and indeed, when applied to the JAGP, yields a binding energy in fair agreement with the value of −5.0 kcal/mol reported by experiments and other theoretical works. The minimum position, the curvature, and the asymptotic behavior of the dispersion curve are well reproduced both at the variational and the LRDMC level. Moreover, thanks to the simplicity and the accuracy of our variational approach, we are able to dissect the various contributions to the binding energy of the water dimer in a systematic and controlled way. This is achieved by appropriately switching off determinantal and Jastrow variational terms in the JAGP. Within this scheme, we estimate that the dispersive van der Waals contribution to the electron correlation is substantial and amounts to 1.5(0.2) kcal/mol, this value being comparable with the intermolecular covalent energy that we find to be 1.1(0.2) kcal/mol. The present Quantum Monte Carlo approach based on the JAGP wave function is revealed as a promising tool for the interpretation and the quantitative description of weakly interacting systems, where both dispersive and covalent energy contributions play an important role.

1. Introduction

The hydrogen bond is a fundamental intramolecular and intermolecular interaction determining the properties of a large number of systems from liquids to solids, from biological[1] to inorganic.[2] Hydrogen bond is commonly defined as a local bond in which a hydrogen atom is attached to an electronegative group (the donor) interacting with another nearby electronegative group (the acceptor) not covalently attached. Dissociation energies cover a range of about 2 orders of magnitude, ranging from −0.2 to 40 kcal/mol, the H-bonding arising from the interplay of different types of interactions. Electrostatic forces play the major role in a large number of hydrogen bonds, although charge transfer effects and van der Waals (vdW) interactions are always present.

* Corresponding author e-mail: leonardo.guidoni@uniroma1.it, http://bio.phys.uniroma1.it.

[†] CASPUR.

[⊥] F.S. and L.S. contributed equally to this work.

[‡] International School for Advanced Studied (SISSA/ISAS).

[#] Current address: Department of Chemistry, University of California, Davis, Davis, CA.

[§] Universita di Roma.

[||] NAST Centre.

10.1021/ct800121e CCC: $40.75 © 2008 American Chemical Society
Published on Web 08/06/2008

Water, the most studied H-bonding liquid, represents the prototype of hydrogen bonded systems. The energetics and directionality of water hydrogen bond is a key factor for understanding the anomalous properties of the water phase diagram,[3] the behavior of small water clusters,[4–6] and the role of aqueous environment in a variety of biological systems.[7] The dissociation energy of the isolated water dimer lies in the middle of the hydrogen bond dissociation energy scale, the most stable configuration being associated with a binding energy $D_e^{\mathrm{exp}} = -5.0$ kcal/mol, as extrapolated by experimental data.[8] The partitioning of this energy in different contribution terms is still the subject of a vivid debate. At the equilibrium bonding distance, typically in a range between 2.5 and 3.5 Å, quantum effects become relevant, and a pure electrostatic picture of the interaction is not fully satisfactory. The partial covalent nature of the hydrogen bond has been recently invoked by the first analysis of the Compton profile on ice *Ih*.[9] However, the interpretation of the experimental data has been questioned by several authors[10,11] and further revised.[12] The amount of the intermolecular covalent contribution, if any, to the binding energy is still an open issue. On the other hand, due to the lack of an unambiguous computational protocol it is still not clear how to estimate the van der Waals contribution to the hydrogen bonding. The role of these interactions may also be at the basis of the current drawbacks of empirical force fields in use for large scale simulations.[13,14]

A definition of the intermolecular covalent component of the hydrogen bonding can be drawn using the intuitive picture of a chemical bond introduced by Pauling as the superposition of Lewis' structures.[15] In the simple case of hydrogen bonding in a water dimer $(H_2O)_2$, three mesomeric Lewis structures may be drawn, one of them describing the charge transfer situation $(OH)^-\cdots(H_3O)^+$, that confers partial covalent character to the hydrogen bond. Within a quantitative Valence Bond representation it would be therefore possible to distinguish, in a simple way, the covalent intermolecular energy contribution by the other interaction energy terms.

At the same time, because of the crucial role of electronic correlation, especially for dispersive interactions, high quality electronic structure correlated methods (based on molecular orbital theory) are necessary for a proper quantitative description of a hydrogen bond. New classes of Density Functional Theory (DFT) functionals have been developed in the past years with the aim to describe weakly bound systems avoiding semiempirical approaches. Nevertheless, the highly nonperturbative and nonlocal character of the vdW interactions makes their inclusions difficult in DFT schemes without the resorting to ad hoc empirical parametrizations.[16] Looking for an ab initio method free of empiricism, the Quantum Monte Carlo[17,18] appears as a possible alternative to other more standard quantum chemistry methods such as Configuration Interaction, Möller Plesset perturbation theory, or coupled-cluster (CC).

Recently, a QMC technique based on the resonating valence bond wave function was introduced in ref 19 and further developed later. This approach represents a very efficient implementation of the valence bond Pauling idea, discovered by P.W. Anderson in the field of strongly correlated electrons:[20] the Jastrow Antisymmetrised Geminal Power (JAGP). This wave function has been demonstrated to be effective in describing highly correlated diatomic molecules like the C_2 as well as π-π interacting complexes.[17,19,21] In the present article we present a Variational Monte Carlo (VMC) and Lattice Regularized Diffusion Monte Carlo (LRDMC) study of the water dimer dissociation energy and dispersion curve, using as a variational ansatz the JAGP wave function. An important advantage of the JAGP VMC approach resides in the possibility of dissecting, in a simple way, the energy contributions of the different terms composing the wave function, like dynamical electron correlations and the intermolecular covalent contribution. Dynamical electronic correlations associated with the charge fluctuations and van der Waals interactions are indeed included in the wave function with Jastrow terms, whereas static correlations are described by the resonance of valence bond singlets in the AGP. The amount of binding energy arising from the correlated dynamical charge fluctuations, related to the vdW forces, can be therefore estimated by the evaluation of the energy contributions of the Jastrow factors. On the other side, following the Pauling idea of chemical bonding we can calculate the energy contribution of the intermolecular covalent term and get insight into the covalent nature of the hydrogen bonding mechanism.

2. Computational Methods

2.1. Geometries. For nuclear coordinates of the water monomer we used the experimental equilibrium geometry[22] with an O−H bond length of 0.9572 Å and a H−O−H angle of 104.52°. For the dimer we used the linear configuration with C_s symmetry, oxygen−oxygen distance 2.976 Å,[23,24] and $O_1-H_1\cdots O_2$ angle of 180°. We used the internal geometry of each monomer as the experimental one. For the dispersion curve we simply used the geometries obtained by shifting away the two monomers along the $O_1-H_1\cdots O_2$ binding axis and keeping fixed their relative orientation. Effects of nuclear relaxation upon binding do not affect our estimations, as we verified by calculating the binding energy with the geometry from CCSD(T) calculations.[25]

2.2. Variational Monte Carlo and the JAGP Wave Function. As variational ansatz we use the JAGP wave function introduced in refs 19 and 21. The wave function Ψ_{JAGP} of a system of N electron is defined by the product of a symmetrical Jastrow term J and an antisymmetrical determinantal part Ψ_{AGP}:

$$\Psi_{JAGP}(r_1, ..., r_N) = \Psi_{AGP}(r_1, ..., r_N)J(r_1, ..., r_N) \qquad (1)$$

The determinantal part Ψ_{AGP} is the antisymmetrized product of spin singlets. The pairing function in singlet system without spin polarization is described by

$$\Psi_{AGP} = \hat{A}\left[\Phi\left(\mathbf{r}_1^{\uparrow}, \mathbf{r}_1^{\downarrow}\right)...\Phi\left(\mathbf{r}_{N/2}^{\uparrow}, \mathbf{r}_{N/2}^{\downarrow}\right)\right] \qquad (2)$$

where $\hat{A}$ is the operator that antisymmetrizes the product of $N/2$ geminal singlets $\Phi(\mathbf{r}\uparrow, \mathbf{r}\downarrow) = \psi(\mathbf{r}\uparrow, \mathbf{r}\downarrow)1/\sqrt{2}(|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle)$. The spatial part of the geminals is expanded over an atomic basis set

$$\psi(\mathbf{r}^{\uparrow},\mathbf{r}^{\downarrow})=\sum_{a,b}\psi_{a,b}(\mathbf{r}^{\uparrow},\mathbf{r}^{\downarrow}) \tag{3}$$

$$\psi_{a,b}(\mathbf{r}^{\uparrow},\mathbf{r}^{\downarrow})=\sum_{l,m}\lambda^{a,b}_{l,m}\phi_{a,l}(\mathbf{r}^{\uparrow})\phi_{b,m}(\mathbf{r}^{\downarrow}) \tag{4}$$

where the indexes l, m run over different orbitals centered on nuclei a, b. The Jastrow factor J is further split into one-body, two-body, and three-body terms ($J = J_1J_2J_3$). The J_1 and J_2 terms deal with electron−electron and electron-ion correlation, respectively. The two-body (one body) Jastrow depends only on the relative distance $r_{i,j} = |\mathbf{r}_i - \mathbf{r}_j|$ between each electron pair (i, j) (electron-ion pair) and has been parametrized by a simple function $u(r_{i,j}) = (1-\exp(-br_{i,j}))/2b$ that rapidly converges to a constant when $r_{i,j}$ became large.[17] In this way the large distance behavior of the Jastrow is determined only by the J_3 Jastrow factor, that contains all variational freedom left and, in particular, as we shall see later on, the slowly decaying vdW correlations. Therefore we have chosen to parametrize this important part of our correlated wave function in a systematic and exhaustive way, similarly to what we have done for the AGP contribution:

$$J_3(\mathbf{r}_1, ..., \mathbf{r}_N)=\exp\left(\sum_{i<j}\Phi^J(\mathbf{r}_i,\mathbf{r}_j)\right) \tag{5}$$

$$\Phi^J=\sum_{a,b}\Phi^J_{a,b}\Phi^J_{a,b}(r_i,r_j)=\sum_{l,m}g^{a,b}_{l,m}\phi^J_{a,l}(r_i)\phi^J_{b,m}(r_j) \tag{6}$$

Both the determinantal $\phi_{a,l}$ and Jastrow $\phi^J_{a,l}$ orbitals are expanded on Gaussian basis sets centered on the corresponding nuclear centers a and b. By increasing the atomic basis set one can rapidly reach the "complete basis set limit" because all cusp conditions are satisfied by an appropriate and simple choice of the J_1 (satisfying the electron-ion cusp) and J_2 (satisfying the electron−electron cusp) terms.[17]

All variational parameters, such as the Jastrow parameters and the $\{g\}$ and the $\{\lambda\}$ matrices of eqs 3 and 5 as well as the exponents and the coefficients of the Gaussian orbitals, have been optimized by energy minimization following the methods described in refs 26 and 27.

The oxygen valence-core interaction was described using the recently reported energy-consistent pseudopotentials.[28] A VMC calculation for the dimer system with the larger basis set and with 0.1 mH accuracy was run for about 12 h on eight AMD Opteron 280 CPUs at the CASPUR computer center. Full wave function optimization was about a factor of 4 more time-consuming.

2.3. Diffusion Monte Carlo. A systematic way for improving the quality of a variational wave function is to perform a Diffusion Monte Carlo calculation, filtering the ground-state properties by a diffusion process.[29] Actually, due to the presence of the fermionic problem, the DMC is implemented within the fixed node (FN) approximation,[30] by imposing that the final ground state has the same nodal structure of the trial WF. In this work we use a slightly modified version of the DMC method, the Lattice Regularized Diffusion Monte Carlo (LRDMC). In this method, the continuum Monte Carlo moves are made by discrete finite steps defined by two lattice spaces a and a'. By using an incommensurate ratio a'/a the electronic trajectory fills the entire space, thus avoiding most lattice artifacts. The introduction of the lattice implies that there are a finite number of possible moves during the diffusion process, and this allows one to avoid the locality approximation and to restore the upper bound property of DMC.[31,32] Within this regularization the exact Hamiltonian H is replaced by a lattice regularized one H_a such that $H_a \rightarrow H$ for $a \rightarrow 0$.[31] We used as lattice spaces the values $a = 0.1, 0.2, 0.3, 0.5$ au, and then the energy was extrapolated to zero lattice space.

Since the dipole moment operator does not commute with the Hamiltonian, in LRDMC we evaluated the estimator $\mu = 2\mu_{LRDMC} - \mu_{VMC}$, where μ_{LRDMC} is the LRDMC mixed average value extrapolated to $a = 0$. A LRDMC calculation for the dimer system with the larger basis set, 0.1 mH accuracy and $a = 0.2$, was run for about 12 h on eight AMD Opteron 280 CPUs at the CASPUR computer center.

2.4. DFT Calculations. For the sake of comparison we perform DFT calculations using a plane wave basis set as implemented in the CPMD code.[33] For the exchange and the correlation part of the universal functional we used BLYP generalized gradient corrections[34,35] and the hybrid functional B3LYP.[36] Core electrons were taken into account using norm-conserving Troullier-Martins type pseudopotentials.[37] We also performed calculations with Dispersion-Corrected Atom-Centered Potentials (DCACP)[16] as described in ref 38. The Kohn−Sham orbitals were expanded in plane waves up to a cutoff of 125 Rydberg.

3. Results and Discussion

3.1. Dissociation Energy and Charge Fluctuations. In this section we report our results on the water dimer at the experimental binding distance, and we investigate the influence of different Jastrow terms of the wave function on the dissociation energy. In the pairing determinant the oxygen atoms are described using a Gaussian basis set of $4s4p$ contracted to $[1s2p]$, whereas we have only an uncontracted $1s$ shell for the hydrogen. We verified that the inclusion of a d-wave shell does not affect the binding energy giving only a rigid shift of the total energy of the dimer and the monomer within LRDMC.

On the contrary, more subtle effects have been observed in the structure of the three-body J_3 Jastrow factor. This term includes in the wave function additional dynamical electron correlations and contributes to the proper behavior of the electronic charge distribution. The correct description of the charge correlations reveals crucial for the inclusion of the dispersive contribution to the vdW interactions, being originated by the correlations between charge fluctuations in different spatial regions.[39]

In Table 1 we report the JAGP monomer energy, E_{H_2O}, the dimer energy, $E_{(H_2O)_2}$, and the dissociation energy, D_e, for increasing three-body Jastrow basis sets. The dissociation energy of the water dimer has been calculated simply as $D_e = E_{dimer} - 2E_{monomer}$. As the number of p-wave shells is increased, we observe an improvement of the binding energy, eventually obtaining at the VMC level a value of D_e = 4.5(0.1) kcal/mol.

The reported LRDMC results, extrapolated to the $a = 0$ limit, appear to have a much faster convergence in terms of

Table 1. VMC and LRDMC Energies for the Water Monomer and Dimer (Atomic Units)[a]

3-body Jastrow basis	E_{H_2O}	$E_{(H_2O)_2}$	D_e
		VMC	
2s2p-local[O]1s[H]	−17.2279(1)	−34.4585(2)	−0.0024(4)[−1.5(0.3)]
2s2p[O]1s[H]	−17.2388(2)	−34.4807(5)	−0.0031(7)[−1.9(0.4)]
2s4p[O]1s[H]	−17.24089(5)	−34.4874(1)	−0.0056(2)[−3.5(0.1)]
2s6p[O]1s[H]	−17.24119(8)	−34.4886(1)	−0.0062(4)[−3.9(0.3)]
2s6p[O]1s1p[H]	−17.2435(1)	−34.4940(1)	−0.0071(2)[−4.5(0.1)]
		LRDMC	
2s2p-local[O]1s[H]	−17.2576(2)	−34.5228(2)	−0.0076(3) [−4.8(0.2)]
2s2p[O]1s[H]	−17.2613(1)	−34.5303(1)	−0.0077(2) [−4.8(0.2)]
2s4p[O]1s[H]	−17.2619(1)	−34.5315(1)	−0.0077(2) [−4.8(0.2)]
2s6p[O]1s[H]	−17.2619(1)	−34.5314(1)	−0.0076(2) [−4.8(0.1)]
2s6p[O]1s1p[H]	−17.2620(1)	−34.5318(1)	−0.0078(2) [−4.9(0.1)]

[a] The dissociation energy calculated as $D_e = E_{(H_2O)_2} - 2E_{H_2O}$ is reported in the last column in atomic units and (in square brackets) in kcal/mol.

total and dissociation energies. This is due to the nature of the projection method which accuracy relies only on the nodal surface of the trial wave function. Our results indicate that the VMC optimized nodal surface, and therefore the corresponding DMC energy, is only slightly affected by the basis set extension of the Jastrow factor. On the other hand, at the VMC level, the size of the Jastrow factor is crucial for improving the binding energy of the dimer. This is probably due to the role of the Jastrow in localizing charges and in introducing dynamical correlations. In this regard the role of *p*-wave shells in the binding energy will be discussed later in more details.

The LRDMC results are obtained extrapolating at zero lattice space and give a binding energy of 4.9 ± 0.1 kcal/mol, in good agreement with the high level quantum chemistry calculations: Klopper et al. have reported 4.99 kcal/mol and 5.02(5) kcal/mol, as a basis set limit for MP2 and Coupled Cluster calculations.[25] Recently a QMC study[18] has reported a value of 5.4(1) kcal/mol, obtained without optimization of the orbitals in the determinant, which is directly taken from a B3LYP calculation.

Our LRDMC results are in the range of previously reported all-electron and pseudopotential QMC calculations.[24,18,40] We also agree with experimental results although they suffer uncertainty due to theoretical estimation of the ZPE. Actually when compared with the experimental dimer dissociation energy $D_e^{\exp}$, the difference between the zero point energy (ZPE) of the monomer and the dimer should be also taken into account: $D_e^{\exp} = D_0 - 2ZPE^{monomer} + ZPE^{dimer}$. The experimental energy reported hereafter is therefore corrected by this quantity calculated by theory or estimated by experiments.[24] As pointed out in ref 17 the JAGP wave function is certainly size consistent for the two water monomers, only when the complete basis set limit is reached for the Jastrow factor. This property can be used to check the basis-set accuracy of the three-body Jastrow term. To verify the size consistency of the wave function we calculated the dissociation energy D_e^* by separating the two monomers by a large distance. D_e^* agrees with D_e.

In QMC calculations, correlation functions different from the energy are often very sensitive to the quality of a wave function. We have therefore calculated the monomer and dimer dipole moment μ that can be easily computed at the VMC level and at the LRDMC level using the mixed estimator. In the case of the water monomer both the variational and the LRDMC dipole moments are rather close to the experimental value of 1.855D.[3] The correction introduced by LRDMC is a slight downshift of the extrapolated estimator, $\mu = 1.870(10)$ D, in agreement with other QMC calculations[18] and ab initio methods.[41]

Table 2. Dipole Moment of Water Monomer and Dimer[a]

3B Jastrow basis	μ_{H_2O} [Debye]	$\mu_{2[H_2O]}$ [Debye]
	VMC	
2s2p-local[O]1s[H]	2.116(17)	2.805(20)
2s2p[O]1s[H]	1.935(12)	2.834(23)
2s6p[O]1s[H]	1.880(8)	2.692(14)
2s6p[O]1s1p[H]	1.890(8)	2.597(12)
	Extrapolated	
2s6p[O]1s[H]	1.874(10)	2.648(18)
2s6p[O]1s1p[H]	1.870(10)	2.603(13)

[a] VMC estimates are reported in the top part of the table. Extrapolated values are reported in the bottom part of the table.

We now turn our attention to the role of dynamical correlations included through the Jastrow term. As discussed above the inclusion of *p*-wave orbitals in the J_3 Jastrow term has a significant effect on the binding energy. Similarly, the effect of the J_3 basis set is also visible on the dipole moments in Table 2. One reason for this influence can be attributed to electrostatic interactions, since the 3-body term is important for the charge distribution. Another relevant effect of the J_3 term is the modulation of the van der Waals interactions. Three effects contribute to the van der Waals forces: induction, thermal orientation, and dispersion. The dispersion forces are quantum mechanical effects originating from the interaction between instantaneous dipoles or, using a second order perturbation theory perspective,[42] by the correlated transition of a couple of electrons from occupied to unoccupied states. Given thus two atomic centers *a* and *b* at a large distance the J_3 term in eq 5 can be expanded for the small value of $g_{ij}^{a,b}$ and then applied to a single geminal pair (see eq 3), $\psi_{a,b}(r\uparrow, r\downarrow)$. The result can be viewed as a correlated transition of two electrons located in different atomic centers from occupied orbitals to unoccupied orbitals with higher angular momentum.[42] More generally the effect of the J_3 at a large distance has the same structure of the vdW perturbative term if on each atomic center the basis used for the Jastrow contains odd orbitals with respect to the spatial reflection, namely when the Jastrow basis set

Table 3. VMC Energy and Dipole Moment of the Water Dimer for Different Jastrow J_3 Terms[a]

pairing terms in J_3	$E_{(H_2O)_2}$ (au)	ΔE kcal/mol)	$\mu_{(H_2O)_2}$ [D]
full $\{g^{a,b}_{l,m}\}$ matrix	−34.4940(1)	0.0	2.597(12)
$(p[H])_1$ $(p[H])_2$	−34.49372(9)	+0.2(1)	2.621(12)
$(p[O])_1$ $(p[H])_2$	−34.4938(1)	+0.1(1)	2.623(12)
$(p[H])_1$ $(p[O])_2$	−34.4935(1)	+0.3(1)	2.610(13)
$(p[O])_1$ $(p[O])_2$	−34.4918(1)	+1.4(1)	2.628(12)
intermolecular p−p $g^{a,b}_{l,m} = 0$	−34.4916(3)	+1.5(2)	2.637(13)

[a] The different J_3 are obtained by canceling the *p*−*p* electronic correlation between atomic centers belonging to different molecules. The atomic center of the *p* wave is indicated between square brackets, and the water molecule index is indicated by the pedex (1 or 2). The energy difference ΔE with respect to the complete $g^{a,b}_{l,m}$ matrix, first line, is also reported in kcal/mol.

contains at least *p* wave orbitals. In principle a small vdW contribution can be derived also from high angular momentum orbitals included in the geminal expansion. In this work however, in order to disentangle the genuine dispersive vdW contribution, we have avoided using polarization orbitals in the AGP that, as discussed before, do not affect the binding energy. In this way the instantaneous correlated polarization induced by the J_3 term allows for the inclusion of dispersive vdW interactions in a transparent variational form.

To understand the effects of the J_3 terms on the dissociation energy, we calculated the variational energy of the wave function obtained excluding intermolecular $g^{a,b}_{l,m}$ terms in eq 5 as reported in Table 3. In particular we considered the H−O and O−O contributions in the *p*−*p* channel, and eventually we eliminated all intermolecular terms (last row of Table 3). Data show nonadditivity of the energy loss, as expected by interactions arising from polarization effects.[43] Among the *p*−*p* wave contribution, the oxygen−oxygen channel seems to be the most relevant term in the Jastrow expansion. It is worth noting that the total dipole moment of the dimer depends only weakly on the intermolecular J_3 Jastrow terms, see Table 3. This indicates that the distribution of the electronic charge is not greatly affected by the missing terms. The energy differences are then due to the part of the dynamical correlation involving correlated excitations to *p* states. The energy loss in the binding energy can therefore be attributed within our formalism to dispersive van der Waals interactions.

It is of interest to compare our result to previous calculations based on symmetry-adapted-perturbation-theory (SAPT)[44,45] that estimated the contribution of dispersion forces to the water dimer hydrogen bond. This contribution amounts to about −1.75 kcal/mol as reported in Table 5 of ref 45. Albeit the energy is not partitioned the same way in the two approaches, the assessment given by SAPT is in good agreement with our estimation of −1.5(2) kcal/mol.

3.2. Dispersion Curve. The VMC and LRDMC dispersion curve of the water dimer is reported in Figure 1A. It has been calculated by computing the total dimer energy as a function of the oxygen−oxygen distance without changing the internal geometry and the relative orientation of the monomers. We used the 2s6p[O]1s[H] basis set for the 3-body Jastrow, which, as reported before, guarantees size consistency during the dissociation process at large distances.

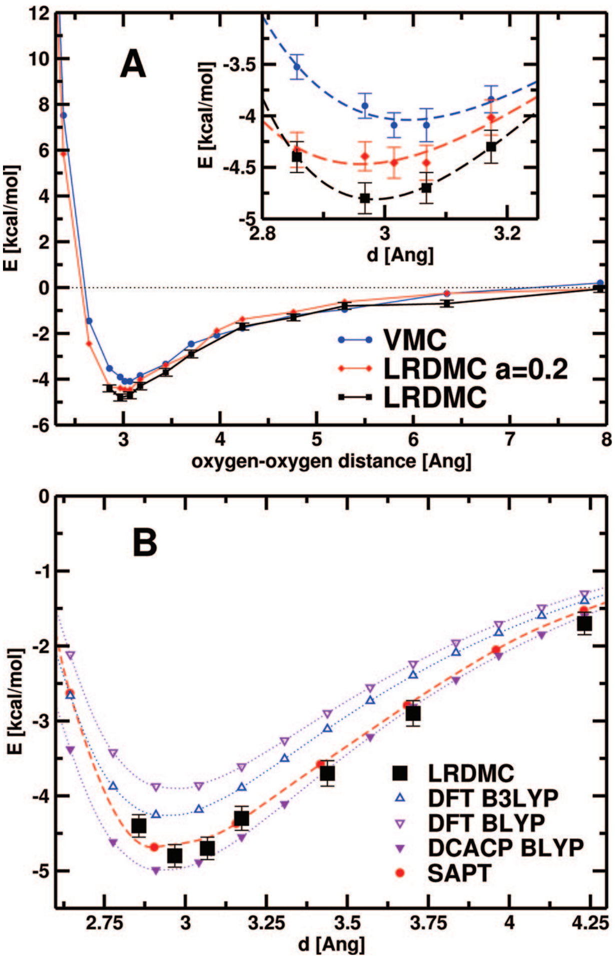


Figure 1. Water dimer dissociation. The total energy of the water dimer is reported as a function of the oxygen−oxygen distance. In Panel A the VMC and LRDMC results are reported. In the inset graph the behavior around the minimum is zoomed in. In Panel B the LRDMC curve is compared with other methods. SAPT values have been taken from ref 46.

The attractive tail of the water−water interaction potential is dominated by a dipole−dipole interaction energy. A polynomial fit for $d \geq 3.5$ Å shows that $E_{2(H_2O)} \sim d^{\alpha}$ with $\alpha = 3.2-3.3$ for the VMC and the extrapolated LRDMC curves, respectively.

The behavior of the dispersion curve at short distances is shown in the inset of Figure 1A, together with a fit performed using a Morse potential. At the VMC level, the minimum of the curve, as obtained by the fitting procedure, is at distance $d = 3.037(4)$ Å, which is slightly shifted with respect to $d = 2.976$ Å reported by experiments.[23] However, it should be noted that, considering the error bars, the curve results to be rather flat around the equilibrium distance. LRDMC with a lattice space $a = 0.2$ au, and the LRDMC extrapolation to zero lattice space $a \rightarrow 0$, improves the location of the equilibrium distance. In this latter case the fitted minimum is at $d = 2.982(1)$ Å, which is very close to the experimental value. In Figure 1B we report a comparison with pure or empirically parametrized Density Functional methods and symmetry-adapted perturbation theory (SAPT).[44,46] Data show that pure BLYP and B3LYP curves underestimate the dissociation energy, whereas calculations performed with

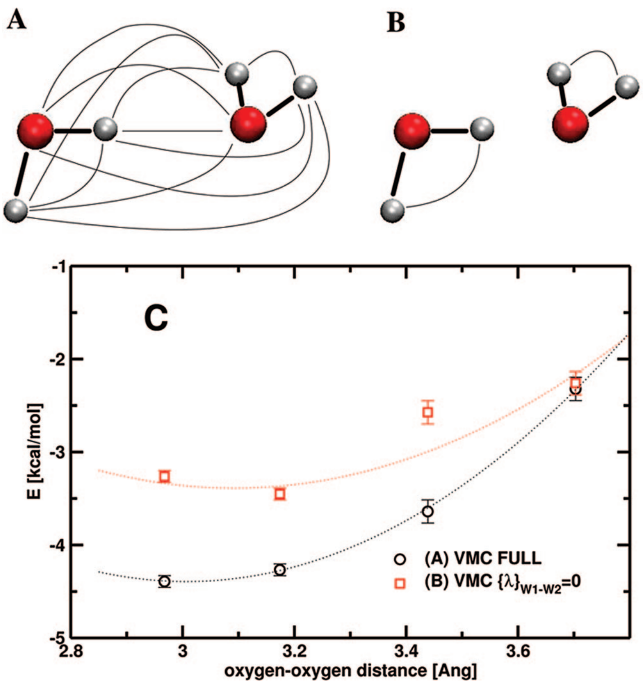


Figure 2. Top panel: Pictorial view of the intermolecular pairing term of the determinant part of the wave function. In Panel A all the intermolecular pairing are drawn. In Panel B we represent the wave function with the intermolecular pairing term set to zero. In the bottom part of the figure the total energy of the water dimer is reported as a function of the oxygen−oxygen distance between the two monomers. Two wave functions are compared, one with all the intermolecular pairing term optimized (black circle) and the other one with intermolecular pairing terms set to zero (red square).

empirically parametrized DCACP pseudopotentials are closer to our LRDMC curve. The SAPT curve is at the top of our results with small discrepancies at very short distance.

3.3. Covalent Contribution to Hydrogen Bonding. In the proximity to the equilibrium distance the interplay between electrostatic and pure quantum effects is expected to be relevant. Although a unique and commonly accepted definition of covalent contribution in a hydrogen bond is still missing, within the formalism of the JAGP wave function, we can define the covalent energy contribution as the energy contribution arising from the intermolecular pairing terms of the RVB determinantal part of the wave function (see eq 3).

The “chemical bond” between the two molecules is indeed due to the superposition of all singlet terms in the geminal expansion that connect two nuclei belonging to different water molecules. This is schematically illustrated in the panel A of Figure 2. In order to evaluate the covalent contribution we proceed as follows. We cut the intermolecular pairing valence bonds in the pairing function, by imposing $\psi_{a,b} = 0$ if a and b belong to different monomers (as sketched in panel B of Figure 2). Then the wave function is reoptimized with the above constraint in order to correctly include the electrostatic effects and the slowly decaying vdW correlations present in our Jastrow factor. In Figure 2C we report, as a function of the oxygen−oxygen distance, the binding energy calculated with the full wave function (circle) and with the wave function lacking the intermolecular valence bond terms (square). The difference between the two curves vanishes as the molecules reach a O−O distance of 3.5 Å. We point out that, by cutting the intermolecular pairing terms, the minimum of the energy dispersion slightly shifts to a larger equilibrium distance.

At the equilibrium distance we found that the contribution of the intermolecular pairing terms, computed at the VMC level, is $\Delta_{inter} = 1.1(0.1)$ kcal/mol, corresponding to about 24% of the computed dimer binding energy. Our estimate of the covalent contribution, defined above, can be compared to what other authors found using different theoretical frameworks and different definition and that is generally referred to as intermolecular charge transfer (CT).[47−50] In the seminal works based on Morokuma decomposition of the binding energy[47,48] the CT contribution to hydrogen bonding is estimated in the range −1.3 to −1.8 kcal/mol, thus about 25% of the total binding energy. A slight smaller contribution, about 11%, resorts from the block-localized wave function approach proposed by Mo et al.[49] They also reported that CT contribution vanishes at about 3.5 Å in good agreement with our finding. It is interesting to observe that the damping of the oscillation at $d \sim 4$ Å of the Fourier Transforms of the Compton profile has been interpreted in ref 12 as a cutoff for the covalent contribution.[9,12] However such an interpretation of the experiments is not fully accepted.[10,11]

4. Conclusions

The understanding of hydrogen bond systems is still a challenge for computational chemistry. Even for small molecular systems the weakness of the interactions and the critical role of electron correlation require the use of affordable correlated quantum chemistry methods. The interplay between interactions different in nature, such as dispersion forces and intermolecular charge transfer, is in many cases crucial for the proper description of the bond properties.

We have shown that the Quantum Monte Carlo method is effective for describing the hydrogen bond between two water molecules. The calculated binding energy matches the experimental value and the estimates from other advanced methodologies. Good agreement with experiments is also achieved for the computed dipole moments. Thanks to the good size scaling properties and the embarrassingly parallelism of QMC algorithms, these methods appear extremely competitive in the context of massive parallel computation.

Moreover, some conceptual advantages rely on the structure of the AGP wave function, a correlated valence bond representation of the electronic system. The AGP formalism gives the possibility to work back on an intuitive picture of localized chemical bonds such as the Pauling’s superposition of Lewis structures. Thanks to the fully correlated structure of the wave function, this picture can be used without compromises in terms of accuracy.

Upon interpretation of the wave function terms, we estimate at the VMC level the covalent contribution to account for 1.1(2) kcal/mol. A similar contribution to the binding energy is given by correlated dipolar vdW fluctuations that account for 1.5(2) kcal/mol.

The quality of our results on the water dimer encourages the application of the method to larger hydrogen bonded systems such as water clusters or small biomolecules. To reduce the computational costs it would be desirable to keep down the number of variational parameters when the size of the system increases. In this respect, different strategies are under investigation.

Acknowledgment. F.S, L.F., and L.G. acknowledge CASPUR for large computational resources and technical and financial support. We thank C. Filippi for providing us with the oxygen pseudopotential before publication. L.G. thanks the Centro Fermi of Rome and for computational resources. L.S., S.S., and L.G. are grateful to INFM and Cineca for grant no. 806 "Progetti Calcolo 2007".

References

(1) Creighton, T. E. *Proteins*; W. H. Freeman and Co.: 1993.

(2) Steiner, T. *Angew. Chem., Int. Ed. Engl.* **2002**, *41*, 49–76.

(3) Franks, F. *Water, a comprehensive treatise*; Plenum: New York, 1972.

(4) Ludwig, R. *Angew. Chem., Int. Ed. Engl.* **2001**, *40*, 1808–1827.

(5) Xantheas, S. S. *Chem. Phys.* **2000**, *258*, 225–231.

(6) Shin, J.; Hammer, N.; Diken, E.; Johnson, M.; Walters, R.; Jaeger, T.; Duncan, M.; Christie, R.; Jordan, K. *Science* **2004**, *304*, 1137–1140.

(7) Ball, P. *Chem. Rev.* **2008**, *108*, 74–108.

(8) Mas, E. M.; Bukowski, R.; K., S.; Groenenboom, G. C.; S., W. P. E.; van der Avoird, A. *J. Chem. Phys.* **2000**, *113*, 6687–6701.

(9) Isaacs, E.; Shukla, A.; Platzman, P.; Hamann, D.; Barbiellini, B.; Tulk, C. *Phys. Rev. Lett.* **1999**, *82*, 600–603.

(10) Ganthy, T.; Staroverov, V.; Koren, P. R.; Davidson, E. *J. Am. Chem. Soc.* **2000**, *122*, 1210–1214.

(11) Romero, A.; Silvestrelli, P.; Parrinello, M. *Phys. Status Solidi B* **2000**, *220*, 703–708.

(12) Barbiellini, B.; Shukla, A. *Phys. Rev. B* **2002**, *66*, 235101.

(13) Cho, C.; Singh, S.; Robinson, G. *J. Chem. Phys.* **1997**, *107*, 7979–7988.

(14) Guillot, B. *J. Mol. Liq.* **2002**, *101*, 219–260.

(15) Bratoz, S. *Adv. Quantum Chem.* **1967**, *3*, 209.

(16) von Lilienfeld, O.; Tavernelli, I.; Rothlisberger, U.; Sebastiani, D. *Phys. Rev. Lett.* **2004**, *93*, 153004.

(17) Sorella, S.; Casula, M.; Rocca, D. *J. Chem. Phys.* **2007**, *127*, 014105.

(18) Gurtubay, I.; Needs, R. *J. Chem. Phys.* **2007**, *127*, 124306.

(19) Casula, M.; Sorella, S. *J. Chem. Phys.* **2003**, *119*, 6500–6511.

(20) Anderson, P. *Science* **1987**, *235*, 1196–1198.

(21) Casula, M.; Attaccalite, C.; Sorella, S. *J. Chem. Phys.* **2004**, *121*, 7110.

(22) Benedict, W.; Gailar, N.; Plyer, E. *J. Chem. Phys.* **1956**, *24*, 1139.

(23) Odutola, J. A.; Dyke, T. R. *J. Chem. Phys.* **1980**, *72*, 5062–5070.

(24) Benedek, N.; Snook, I.; Towler, M.; Needs, R. *J. Chem. Phys.* **2006**, *125*, 104302.

(25) Klopper, W.; van Duijneveldt-van de Rijdt, J.; van Duijneveldt, F. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2227–2234.

(26) Sorella, S. *Phys. Rev. B* **2005**, *71*, 241103.

(27) Umrigar, C.; Toulouse, J.; Filippi, C.; Sorella, S.; Hennig, R. *Phys. Rev. Lett.* **2007**, *98*, 110201.

(28) Burkatzki, M.; Filippi, C.; Dolg, M. *J. Chem. Phys.* **2007**, *126*, 234105.

(29) Foulkes, W.; Mitas, L.; Needs, R.; Rajagopol, G. *Rev. Mod. Phys.* **2001**, *73*, 33.

(30) ten Haaf, D. F. B.; van Bemmel, H. J. M.; van Leeuwen, J. M. J.; van Saarloos, W.; Ceperley, D. M. *Phys. Rev. Lett.* **1994**, *72*, 2442–2445.

(31) Casula, M.; Filippi, C.; Sorella, S. *Phys. Rev. Lett.* **2005**, *95*, 100201–100204.

(32) Casula, M. *Phys. Rev. B* **2006**, *74*, 161102(R)

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

(34) Becke, A. *Phys. Rev. A* **1988**, *38*, 3098–3100.

(35) Lee, C.; Yang, W.; Parr, R. *Phys. Rev. B* **1988**, *37*, 785–789.

(36) Becke, A. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

(37) Troullier, N.; Martins, J. *Phys. Rev. B* **1991**, *43*, 1993–2006.

(38) Lin, I.; Coutinho-Neto, M.; Felsenheimer, C.; von Lilienfeld, O.; Tavernelli, I.; Rothlisberger, U. *Phys. Rev. B* **2007**, *75*, 205131.

(39) Dobson, J.; Wang, J.; Dinte, B.; McLennan, K.; Le, H. *Int. J. Quantum Chem.* **2005**, *101*, 579–598.

(40) Diedrich, C.; Luchow, A.; Grimme, S. *J. Chem. Phys.* **2005**, *123*, 184106.

(41) Coutinho, K.; Guedes, R.; Cabral, B.; Canuto, S. *Chem. Phys. Lett.* **2003**, *369*, 345–353.

(42) Cohen-Tannoudji, C.; Diu, B.; Lalo, F. *Quantum Mechanics*; Wiley-Interscience: 1977; *Vol. II.*

(43) Lifshitz, E. M.; Landau, L. D. *Quantum Mechanics: non-relativistic theory*; Butterworth-Heinemann: 1981; *Vol. III.*

(44) Mas, E. M.; K, S.; Bukowski, R.; Jeziorski, B. *J. Chem. Phys.* **1997**, *107*, 4207.

(45) Misquitta, A. J.; Podeszwa, R.; Jeziorski, B.; Szalewicz, K. *J. Chem. Phys.* **2005**, *123*, 214103.

(46) Wu, X.; Vargas, M.; Nayak, S.; Lotrich, V.; Scoles, G. *J. Chem. Phys.* **2001**, *115*, 8748–8757.

(47) Morokuma, K. *Acc. Chem. Rev.* **1977**, *10*, 294–300.

(48) Singh, U. C.; Kollman, P. *J. Chem. Phys.* **1985**, *83*, 4033–4040.

(49) Mo, Y.; Gao, J.; Peyerimhoff, S. *J. Chem. Phys.* **2000**, *112*, 5530–5538.

(50) Iwata, S.; Nagata, T. *Theor. Chem. Acc.* **2007**, *117*, 137–144.

CT800121E

JCTC Journal of Chemical Theory and Computation

Beyond Point Charges: Dynamic Polarization from Neural Net Predicted Multipole Moments

Michael G. Darley, Chris M. Handley, and Paul L. A. Popelier*

Manchester Interdisciplinary Biocentre (MIB), 131 Princess Street, Manchester M1 7DN, United Kingdom

Received May 15, 2008

Abstract: Intramolecular polarization is the change to the electron density of a given atom upon variation in the positions of the neighboring atoms. We express the electron density in terms of multipole moments. Using glycine and *N*-methylacetamide (NMA) as pilot systems, we show that neural networks can capture the change in electron density due to polarization. After training, modestly sized neural networks successfully predict the atomic multipole moments from the nuclear positions of all atoms in the molecule. Accurate electrostatic energies between two atoms can be then obtained via a multipole expansion, inclusive of polarization effects. As a result polarization is successfully modeled at short-range and without an explicit polarizability tensor. This approach puts charge transfer and multipolar polarization on a common footing. The polarization procedure is formulated within the context of quantum chemical topology (QCT). Nonbonded atom−atom interactions in glycine cover an energy range of 948 kJ mol^{-1}, with an average energy difference between true and predicted energy of 0.2 kJ mol^{-1}, the largest difference being just under 1 kJ mol^{-1}. Very similar energy differences are found for NMA, which spans a range of 281 kJ mol^{-1}. The current proof-of-concept enables the construction of a new protein force field that incorporates electron density fragments that dynamically respond to their fluctuating environment.

1. Introduction

For large systems, ab initio calculations quickly become computationally expensive and even prohibitive when simulations need to be carried out. Nonetheless, force fields dramatically accelerate simulations or even enable them. This is often achieved at the expense of accuracy. The associated potentials must be quantitatively accurate,[1,2] however, if one wishes to study problems such as molecular recognition,[3,4] polymorphism,[5] and protein conformation.[6−8]

Force fields suffer from a reliance on parameter fitting and various a priori simplifications, which restrict the transferability of the potentials. A popular simplification in force field design is the assignment of point charges to atoms. Here, the atomic charge density, which is ultimately responsible for the electrostatic interaction, is boldly assumed to be spherical and replaced by a monopole (moment). Another popular simplification occurs in the treatment of polarization where atomic charges are deliberately and permanently enhanced. Alternatively, a dynamic response to a fluctuating external field is modeled by atomic charges acquiring a companion charge, attached to the atom by a fictitious spring. The importance of accurately modeling polarization is widely accepted, as illustrated by a special issue of this journal, published in 2007, dedicated to this critical problem.[9] Furthermore, in 2004 Gresh et al. demonstrated[10] the need for ab initio distributed multipoles and anisotropic distributed polarizabilities in a study on tetrapeptides. Subsequently, a high level MP2 study[11] on conformations of amino acids highlighted the importance of polarizability (as well as multipole moments) as important factors in the design of new force fields. This finding is echoed in recent ab initio conformational work[12] on seven pilot molecules.

Despite dramatic advances in computer hardware, most force fields still adopt the aforementioned simplifications. For example, CHARMM,[13] AMBER,[14] GROMOS,[15] and OPLS[16] treat atoms as nonpolarizable point charges. In

* Corresponding author e-mail: pla@manchester.ac.uk.

10.1021/ct800166r CCC: $40.75 © 2008 American Chemical Society
Published on Web 08/07/2008

general, the assignment of atomic charges is not straightforward, and there are many ways that molecular charge distribution can be separated into atomic components. Most modern force fields rely on fitting individual charges to a molecular electrostatic potential (MEP).[17−20] In their careful and deep analysis of this fitting problem Chirlian and Francl showed[21,22] from singular value decompositions of the least-squares matrices that statistically valid charges cannot be assigned to all the atoms in a given molecule. As a result, atoms in similar chemical environments can be given different fitted charges. Ignoring this careful mathematical analysis, arbitrary penalty functions were introduced[23] on the spurious physical grounds that a "buried" atom would somehow contribute less to a perfectly additive electrostatic potential. Another problem arises if we consider different conformations of the same molecule. Different conformations give different MEPs, and hence different charges may be fitted for the same atoms. Furthermore, the fitting procedure does not account for the internal polarization as the charge density changes with conformation. Most MEP fitting procedures take the mean charge for each atom from a number of conformations[24] or fix the charge of a particular atom type to be the same.[23]

The point charge representation is also inadequate for explaining the relative stability of crystal polymorphs of organic molecules or the structural motifs important for molecular recognition.[3,5] These cases are dependent on the strength and the directionality of intermolecular interactions, the directionality being due to the anisotropic distribution of charge. Because of the inability to represent this anisotropic distribution, many point charge models introduce "off-atom" sites. A prime example of this is the placing of off-atom sites at the "lone-pair" locations about oxygen atoms.[25−28]

A more accurate description of the charge distribution uses multipole moments, which can be regarded as the original coefficients of the series expansion that describes the electrostatic potential. Although elaborate compared to point charges, multipole moments can be expressed compactly (and irreducibly) in terms of spherical harmonics.[29] Multipole moments can be determined in a number of ways. Distributed multipole analysis (DMA)[30] determines the multipole moments from the wave function of a molecule by analyzing the overlap between Gaussian functions. DMA moments are employed in force fields such as AMOEBA[31] or the effective fragment potential (EFP) method.[32] Alternatively, the partitioning scheme of Vigné-Maeder and Claverie[33] is used within the "sum of interactions between fragments ab initio" (SIBFA) potential.[34,35] As a further alternative, the Gaussian electrostatic model (GEM)[36,37] approach, which has evolved from SIBFA, uses a full charge distribution based on a density fitting scheme rather than multipole moments. Another way of partitioning the charge density is by means of Wannier functions.[38,39] They were employed to determine multipole moments in simulations, but they were not directly involved in determining the electrostatic interaction.[40] Moments derived by this method have been implemented in simulations of biomolecules.[41]

In this paper we work with the partitioning method of the quantum theory of "Atoms in Molecules",[42,43] which we consider as part of the quantum chemical topology (QCT) approach, a name we justified in ref 44. The QCT approach partitions the electron density into *finite* topological atoms, which exist in real space. The corresponding atomic multipole moments are then obtained by integration of the appropriate density over the atomic volumes. Note that the DMA method allows multipoles to be centered on non-nuclear sites, such as in the middle of a chemical bond. Naively one would think that this flexibility is missing in the QCT approach in which the multipole expansion site coincides with a nucleus, unless there is a non-nuclear attractor with its own multipole moments. However, thanks to a clear distinction between distribution and partitioning[45] one can shift QCT *atomic* multipole moments to a site away from the nucleus, such as the bond midpoint.[45] Alternative chemically intuitive distributed electrostatic moments can be obtained,[46] using the topology of ELF,[47] a prime component of the QCT approach.

QCT moments have been successfully used in the simulation of liquid HF[48] and liquid water.[49,50] These results were preceded by work that corroborated the success of QCT multipole moments in the reproduction of the electrostatic potential[51] and electrostatic or Coulomb atom−atom interaction[52,53] and the prediction of structure of nucleic acid−base pairs.[54] In this systematic work on QCT potentials, the important issue of the convergence of the multipole expansion was featured strongly, leading to solutions that increased the convergence radius[55,56] or accelerated convergence.[45] We also showed that 1−3 and 1−4 interactions can be expressed as a convergent multipole expansion,[57] putting these interactions on a par with nonbonded interactions ($1 - n$, $n > 4$). Even exchange energy can be favorably expanded in terms of exchange moments, but their transferability remains elusive.[58] A separate study[59] determined the lowest rank necessary to achieve a preset error in atom−atom interaction energy.

In the aforementioned work the atomic charges were kept constant with respect to any changes in the orientation or conformation of the molecules. However, this is not representative of reality. The distribution of electron density shifts in response to changes in the intramolecular and intermolecular interactions as orientations and configurations change. Polarization can be modeled by distributed polarizabilities,[60,61] pioneered within the QCT context by Ángyán et al.,[62] and later applied to study[63] field-induced charge transfer along the hydrogen bonding in the water dimer. Since then we have taken a drastically different route to model polarization, which does not introduce an explicit polarizability tensor. Instead, the polarizability is implicit (and still fully anisotropic), in that a machine learning technique is trained to predict atomic multipole moments from the positions of all other atoms in the molecule. The prediction formulas are analytical and nonlinear. They are stored within a trained neural network and can be retrieved, in principle, although this is typically not done in the neural network literature. Of course, in its convoluted analytical shape, the trained neural net can be exported into a molecular dynamics package, for example. Second, it should be emphasized that the current approach abandons upfront the picture of an external electric field that causes the polarization. Instead, we embrace the more general picture that electron density fragments change in response to a change of nuclear positions, which cause the field. This field, which applies at very short-range and which can be most inhomogeneous (anisotropic), exerts its influence through the ab initio calculations that generate the electron density in the first place. We only need to focus on the result of the field, whether intra- or intermolecular in nature, without approximating it in any way. We believe that this approach automatically takes care of the nonadditivity of polarization. Furthermore, if atoms are taken out of a sufficiently large environment, their electron density has already been adjusted. Hence, in principle there is no need for an iterative scheme that adjusts the atom and its neighbors using an explicit polarizability tensor.

We published the first example and proof-of-concept of the idea of implicit neural net polarization by means of a molecular dynamics simulation[64] of a polarizable hydrogen fluoride dimer. The machine learning method consists of standard backpropagation (artificial) neural networks (NNs). Fluctuating QCT multipole moments, which expressed the intermolecular polarization, were successfully modeled by NNs. The current paper presents the first example of this novel technique applied to intramolecular polarization, illustrated by the molecules glycine and *N*-methylacetamide (NMA).

NNs mimic the way a brain functions by using an array of interconnected units that pass information between themselves to recognize complex patterns. Through the modification of the strength of the connections between these units, NNs are able to learn functions. NNs have been applied to a number of systems and are able to represent them without any prior knowledge about the form of the potentials which govern such systems.[65] Instead, NNs are able to learn the true underlying function from a set of pregenerated data. NNs have been used for modeling the Al^{3+} system,[66] a hydrocarbon potential,[67] silicon,[68] the H_3^+ ion,[67] and water.[69,70]

Here we employ NNs to learn the relationship between the atomic multipole moments and the nuclear configuration of its environment (which is the whole molecule for both glycine and NMA). This makes possible a dynamic representation of the electron density, able to react to changes in the local environment. This eliminates the need for ad hoc corrections[32,71] that account for the overlap of electron densities,[72] an inevitable problem when using fixed densities. Additionally, this NN approach allows us to incorporate the effect of charge transfer into the prediction of the multipole moments. Polarization and charge transfer are now treated on a par with traditional electrostatic interactions, as a single dynamic electrostatic term.

2. Background

2.1. Polarization. The effect of polarization is not negligible and is often quoted as accounting for ∼15% of the total interaction energy,[73,74] with Yu and van Gunsteren quoting a range of 10−50% of the total interaction energy.[75] Including polarization allows for the fluctuating anisotropic nature of atomic electron densities to be correctly modeled, which is important for the calculation of accurate interaction densities and the accurate determination of the relative stabilities of molecular conformations,[76] polymorphs, and molecular clusters.[77] The importance of polarization can be highlighted by the dipole enhancement of water molecules transferred from the gas to the bulk phase.[78−80]

There are a number of methods for adding explicit polarization to potentials, but the three main methods are polarizable point dipoles, fluctuating charge models, and Drude oscillators.

Polarizable point dipoles are represented by two point charges. These dipoles then interact via a tensor just as static multipole moments do. The magnitude of the dipole is a response to an external field and is calculated iteratively as molecules respond to mutual changes in charge distribution. However, there is a danger of a “polarization catastrophe”, which results from the polarization amplifying itself, causing the interaction energy to become infinite. In the AMOEBA model this is prevented by using a Thole-type damping function.[31,81−85] Within the SIBFA model, point dipoles are distributed about the molecules and are placed at lone pair centers and bond barycenters rather than nuclei.[86] This procedure is analogous to the placement of point dipoles in the EFP model.[87] In contrast to AMOEBA, the SIBFA model tackles the polarization catastrophe by introducing a Gaussian screening of the field. Piquemal et al.[88] re-evaluated the SIBFA polarization compared to its counterparts. Overpolarization at short distances can also be overcome by a new method parametrizing a polarizable potential based on Car−Parrinello simulations.[89]

Fluctuating charge models,[90] or charge equalization, iteratively modify atomic charges in response to an external field. The TIP4P-FQ model of water[91,92] is an early application of this method. Its main disadvantage is that it only accounts for isotropic polarization.

Drude oscillators,[93] or charge-on-spring models, represent polarization by two point charges that are tethered by a harmonic spring potential. One of the charges is located at a fixed position while the other is free to move in response to an external field. This method allows for an anisotropic polarization response.[94−96]

Another non-negligible energy component that is not explicitly represented in nonpolarizable models is charge transfer.[97,98] Charge transfer is the partial transfer of charge from a donor to an acceptor molecule, or between atoms, and so affects the electrostatic interactions of the atoms. Thus charge transfer is a more extreme case of polarization.[99] To the best of our knowledge, the only polarization scheme that is currently able to incorporate charge transfer is the FQ approach, which allows the atomic charges to change in response to the chemical environment. Other models account for the effect of charge transfer by including a further component in the potential.[34−36,100] There is an explicit representation of the charge transfer term in the context of SIBFA going back to 1982.[101] Chen and Martínez made the important observation that the electronegativities that determine the charge transfer are geometry dependent.[99]

2.2. Quantum Chemical Topology and Electrostatic Interaction. Topological atoms are naturally carved out by gradient paths in the electron density. These paths of steepest ascent typically originate at infinity and terminate at a nucleus. They create surfaces that bound the electron density that is then naturally allocated to each atom. An important feature of topological atoms is that they do not overlap and that they exhaust space (i.e., leave no gaps between them). Atomic multipole moments are obtained by integrating the appropriate property density over the atomic volume. The property density is the total charge density multiplied by a regular spherical harmonic,[102] for example, $^1/_2(3z^2 - r^2)$ for one of the five quadrupole moments.

The moments of atom A and B are designated as $Q_{l_{\mathrm{A}}}m_{\mathrm{A}}(\Omega_{\mathrm{A}})$ and $Q_{l_{\mathrm{B}}}m_{\mathrm{B}}(\Omega_{\mathrm{B}})$, respectively, where the index l refers to the rank of the multipole moment and m to the component. Each atom has its own local axis system, centered on its nucleus. The Coulomb interaction between two atoms is then given by[52]

$$E^{\mathrm{AB}} = \sum_{l_{\mathrm{A}}=0}^{\infty} \sum_{l_{\mathrm{B}}=0}^{\infty} \sum_{m_{\mathrm{A}}=-l_{\mathrm{A}}}^{l_{\mathrm{A}}} \sum_{m_{\mathrm{B}}=-l_{\mathrm{B}}}^{l_{\mathrm{B}}} T_{l_{\mathrm{A}}m_{\mathrm{A}}l_{\mathrm{B}}m_{\mathrm{B}}}(\mathbf{R})\, Q_{l_{\mathrm{A}}m_{\mathrm{A}}}(\Omega_{\mathrm{A}})\, Q_{l_{\mathrm{B}}m_{\mathrm{B}}}(\Omega_{\mathrm{B}}) \quad (1)$$

where $T(\mathbf{R})$ is the interaction tensor and $\mathbf{R}$ the vector linking the nuclear positions of the respective atoms (i.e., the origins of their local frames). The terms of eq 1 can be collected according to powers of $R = |\mathbf{R}|$ by means of a rank called L, which is defined as $l_{\mathrm{A}} + l_{\mathrm{B}} + 1$. For example, R^{-3} dependence is made up of interactions between two dipole moments and between a monopole moment and a quadrupole moment. The convergence of the multipole expansion can be monitored against a varying rank L. Using Hättig's recurrence formula[103] for the interaction tensor we can generate expansions up to arbitrarily high rank. The exact interaction energy can be obtained via a six-dimensional (6D) integration over the two participating atoms,

$$E^{\mathrm{AB}} = \int_{\Omega_{\mathrm{A}}} \mathrm{d}\mathbf{r}_{\mathrm{A}} \int_{\Omega_{\mathrm{B}}} \mathrm{d}\mathbf{r}_{\mathrm{B}} \frac{\rho_{\mathrm{tot}}(\mathbf{r}_{\mathrm{A}})\, \rho_{\mathrm{tot}}(\mathbf{r}_{\mathrm{B}})}{r_{\mathrm{AB}}} \quad (2)$$

where r_{AB} is the distance between two infinitesimally small charge elements and ρ_{tot} is the total charge density.

2.3. Neural Networks. NNs, which have been the subject of interest[104−106] for many years, are a well-researched and popular example of a machine learning technique. We can only give a brief account here and refer to the Appendix and the citations above for further detail. A NN is an array of connected *nodes*, which pass information between themselves. Each node receives a number of inputs and sends an output. The nodes sum their inputs, which are individually multiplied by the relevant weights, and pass this sum through a *transfer function,* which gives the output. It is the alteration of these weights that allows a NN to learn functions.

The architecture of a network is defined by the number of hidden layers and the number of nodes in the input, output, and hidden layers. The hidden layer contains *hidden nodes*, so-called because one does not have direct access to their outputs for the purpose of training. Hence, they must develop their own representation of the input.[105] The hidden layer enables the network to learn complex tasks by extracting progressively more meaningful features from the input patterns. One of the simplest is a single hidden layer feedforward network (see Figure 1). In a feedforward network the nodes only pass information to the next layer and not back to a previous layer. To learn a mapping between input and output patterns, the NN is presented with training data. During the learning process the NN errors in predicting (i.e., reproducing) the values of the training data are used to alter the weights. This is called the *backpropagation of errors* method. More details are given in the Appendix. The process is repeated for every example in the training set before beginning again, with each full pass of the training set being called an *epoch*.

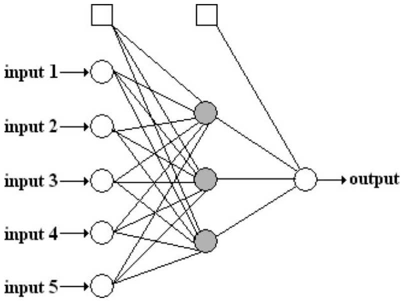


Figure 1. Diagram of a feedforward NN with one hidden layer. The gray circle is a hidden node, a square is a bias, and each connection represents an adjustable weight. In this work, the output is a multipole moment of a given atom and the inputs are the coordinates of neighboring atoms.

Each neuron receives a number (p) of inputs. Each input (signal) is associated with a weight, which can be positive (excitatory) or negative (inhibitory). The activation of neuron k, denoted by a_k, is then defined as the sum of the products of the input and the corresponding weights w_{kj}, or

$$a_k = \sum_{j=0}^{p} w_{kj} x_j \quad (3)$$

Note that, in general, the weight w_{kj} is associated with the connection between neuron j and neuron k. The actual output of a neuron depends on its activation, which has to exceed a given threshold θ for the neuron to fire. Here, the sigmoid function σ acts as a nonlinear transfer function determining how a neuron's output depends on its activation. It is defined in eq 4,

$$y = \sigma(a) = \frac{1}{1 + \exp[-(a - \theta)/\rho]} \quad (4)$$

where ρ controls the shape of the sigmoid. In this work, the output is a multipole moment of a given atom and the inputs are coordinates of neighboring atoms.

To achieve the optimal NN for the prediction multipole moments the architecture of the NN is modified, in this case, only by varying the number of hidden nodes. The net's performance can also be improved by tuning two training parameters, the learning rate and the momentum.[105] Before training the input data must be standardized (see Appendix), that is, transformed to dimensionless data that have a mean value of zero and a standard deviation of one. The data are then transformed to lie in the interval [0, 1] (i.e., normalized) via eq 5,

$$x_n = \frac{(x - x_{\min})}{(x_{\max} - x_{\min})} \tag{5}$$

where x is a standardized input, x_n is a normalized input, $x_{\min}$ and $x_{\max}$ are the minimum and maximum values, respectively, and the subscript n refers to the normalization.

When training the NN we must consider its performance using statistical measures. We make use of the r^2 correlation coefficient, which measures the linear relationship between the predicted output and the desired output, defined in eq 6,

$$r^2 = 1 - \left[\frac{\sum_{j=1}^{N}(a_j - b_j)^2}{\sum_{j=1}^{N}\left(a_j - \left(\frac{1}{N}\sum_{j=1}^{N} a_j\right)\right)^2}\right] \tag{6}$$

where a_j is the target output, b_j the predicted output, and N the number of training examples. The denominator is reminiscent of a standard deviation. An analogous test is performed using a test set of data, unseen by the network during its training. Here the test set has the same size as the training set. The corresponding correlation coefficient is referred to as q^2.

When training a network one should be concerned about generalization, which covers both overfitting and overtraining. Overfitting means that a net has too much flexibility and thereby inappropriately accommodates all the noise and intricacies in the data without regard for the underlying trends.[105] This undesirable effect can be detected if q^2 turns out to be much lower than r^2. Overtraining can occur if the training continues for too many epochs. It is then possible that the net offers no predictive ability for examples that it was not trained with. Rather than monitoring the number of epochs, overfitting can again be detected by q^2 being much lower than r^2. In summary, a properly generalized net (i.e., not suffering from overfitting or overtraining) has an r^2/q^2 ratio close to unity.

3. Computational Details

To properly calculate the interaction energies we must orient the anisotropic atoms correctly. For a given atom we define an atomic local frame (ALF), which is determined by the connectivity of this atom. Next, the multipole moments of this atom are defined with respect to its ALF. This orientation procedure is adopted during the preparation of the training data as well as during the deployment of the NNs as a way to align the moments by the neural nets in an energy minimization or a simulation. The ALF defines the rotation of the predicted moments into the global frame in which the molecule resides. The NNs knowledge of the local chemical environment is stored in the ALF. In summary, the NN both gains information from the ALF and predicts moments within this ALF, in a unified and consistent way.

The method for determining an ALF uses the Cahn−Ingold−Prelog rules for determining the absolute configuration of a chiral center. Note that we do *not* use these rules to determine the chirality of our atoms; we only adopt that part of the rules that ranks groups (attached to a given atom) according to priority. The central atom, whose moments are being predicted, defines the origin of the ALF. The x-axis is determined by the atom with the highest atomic number neighboring the central atom. The xy-plane is then determined by the neighboring atom with the next highest atomic number. In the case of ambiguity, the Cahn−Ingold−Prelog rules are able to decide which atom acquires priority by inspection of the atomic number of the next or more distant atoms. For terminal atoms the process is simpler because the atom has only one connected neighbor, and hence this atom defines the x-axis. To define the xy-plane we inspect the atoms connected to the x-axis atom and, according to the above rules, again determine the atom with the highest atomic number. This atom defines the xy plane. This procedure is illustrated in Figure 2.

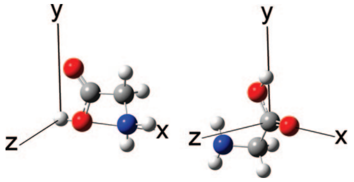


Figure 2. Glycine molecule rotated into two atomic local frames. The left orientation is an example of an atomic local frame for a terminal atom. The right orientation is an example of an atomic local frame for a nonterminal atom.

Once the ALF is defined, the training data are generated. The molecule is geometry- optimized to a local minimum and the corresponding frequencies of the normal modes of vibration are calculated. Then the molecule is distorted randomly along the normal modes of vibration. Each normal mode is distorted in turn by a random amount, according to a randomly assigned sign of displacement, either positive or negative. However, before the displacement is applied we ensure that the total amount of energy does not exceed the imposed limit of 20 kJ mol^{-1}. If so, the distortion is performed again. Figure 3 illustrates a set of geometries for glycine and NMA, and their atomic labeling schemes used throughout this article.

The number of molecular configurations (or geometries) needed for the training of the NN depends upon its architecture, in particular, its number of weights. As a general rule, approximately 10 training examples (i.e., molecular geometries) are required for every weight. Equation 7 provides the number of weights, N_{weight}, for a given architecture with N_{input} input nodes, N_{hidden} hidden nodes, and N_{output} output nodes,

$$N_{\text{weight}} = (N_{\text{input}} + 1)N_{\text{hidden}} + (N_{\text{hidden}} + 1)N_{\text{output}} \tag{7}$$

One adds 1 before multiplication to account for a bias weight. Note that bias nodes only feed connections forward; they do not accept connections. In this study, $N_{\text{input}} = 24$ because glycine has $3N - 6 = 3 \times 10 - 6 = 24$ internal degrees of freedom, and all these degrees are taken to influence the multipole moments of a given atom. Invariably, each multipole moment constitutes the single output for each

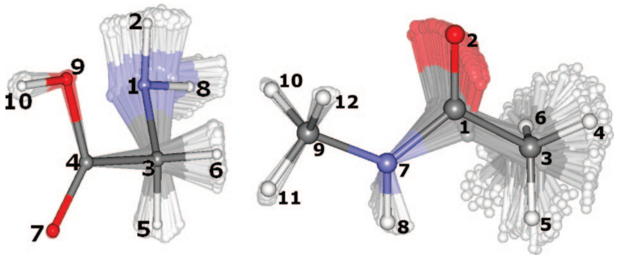


Figure 3. Representative sets of glycine (left) and *N*-methylacetamide (right) molecules distorted along their normal modes of vibration. The reference (local) minimum energy geometry is shown in bold. In the glycine set C_4 defines the origin, the O_7 the *x*-axis, and O_9 the *xy*-plane. For the *N*-methylacetamide set C_9 defines the origin, the N_7 the *x*-axis, and H_{10} the *xy*-plane. In the distorted structures only the atom that defines the origin is in exactly the same position in each structure, while the atoms that define the *x*-axis all have slightly different positions on the *x*-axis due to the random distortion applied to the bond. Similarly, the atoms which define the *xy*-plane all have different positions in the *xy*-plane.

NN, hence, N_{output}=1. Equation 7 then specializes to $N_{\text{weight}} = 26N_{\text{hidden}} + 1$. Because the maximum number of hidden nodes in this work is 10, the maximum number of weights is 261 and hence an adequate number of training examples is 2610. We generated enough training examples to accommodate the largest NN architecture we use.

Multipole moments up to the quadrupole are predicted using the NNs. Higher order moments are not predicted but instead are taken as fixed from the reference geometry (which is local energy minimum). Each moment is predicted by a single NN so there are nine NNs per atom: one for the monopole, three for the dipole, and five for the quadrupole moment.

The NN used to predict the atomic moments is selected using the r^2/q^2 ratio and the NN parameters. The latter refer to the number of hidden nodes, the momentum, and the learning rate. The NN that has an r^2/q^2 ratio closest to unity is selected regardless of that NN's architecture. However, if two architectures have the same r^2/q^2 ratio, then the NN with the smallest number of nodes is selected. If the NNs have the same r^2/q^2 ratio and number of hidden nodes, then the one with the lowest momentum is selected.

The program GAUSSIAN03[107] geometry-optimized glycine and NMA, and calculated the frequencies for the normal modes of vibration. The frequencies are then used to guide the generation of the distorted structures. An energy limit of 20 kJ mol^{-1} is imposed, and the molecule is randomly distorted along the normal modes. This maximum distortion energy is set so that the generated geometries are plausible. The wave functions of each of the distorted geometries are generated and are then used by the program MORPHY[108] to compute the atomic multipole moments.

For this pilot study, all optimizations, frequency, and wave function calculations were performed at B3LYP/6-311+G(2d,p) level. The proposed method is purely based on the electron density and hence does not depend on the details of how this density was generated.

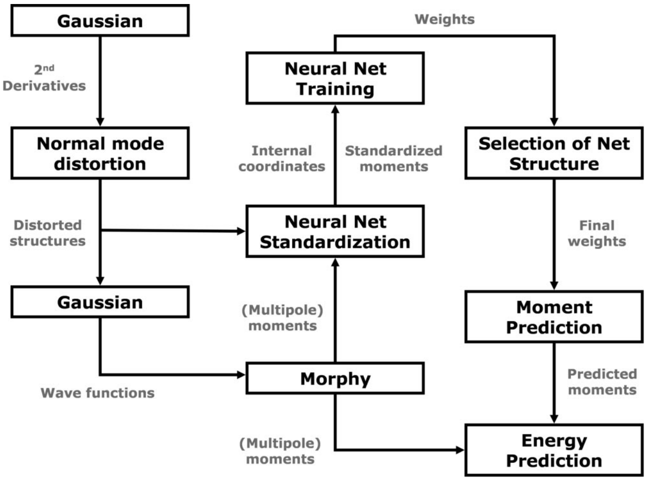


Figure 4. Steps used to create the data for NN training and moment prediction. Black text in boxes represents programs that produce output used by other programs. Grey text represents the output of a program step that is used as input in following programs.

We generated 2610 distorted geometries for each atom with 1305 geometries for NN training and 1305 as a test set. A further 50 distorted molecules were kept in the global frame (i.e., not rotated into an ALF). This set served as a validation set. One could assess the NN's performance by means of a correlation coefficient along the lines of eq 6. However, it is more informative to invoke the interaction energy between a given atom and another atom to assess the NN. For the validation set the *true* moments are calculated from the wave functions and then used to calculate the interaction energy between a subset of all possible atom pairs in glycine. This subset consists of 23 pairs, which is the total number of pairs (45 = 10(10-1)/2) minus all 9 bonded (1-2) interactions and all 13 valence angle (1-3) interactions. In other words, only the remaining 23 interactions of the type 1-4 and 1-*n* ($n \geq 4$) were considered. For glycine, these 23 *true* interaction energies are compared with the interaction energies calculated using the *NN predicted* moments. For NMA, this comparison involved 37 atom pairs.

Figure 4 outlines the data generation procedure. The boxes represent steps that produce output used in the following steps. The first step is the optimization of the molecule using GAUSSIAN, which also computes the second derivatives of the optimized structure. The latter are used to generate the distorted structures in the second step. A single-point calculation is performed for each distorted structure and the corresponding wave function file passed on the program MORPHY. The multipole moments of all atoms for each distorted structure are calculated by MORPHY. The Cartesian coordinates are converted into internal coordinates, which are then standardized and transformed (according to eq 5) to attain a value between 0 and 1. The same transformations are applied to the multipole moments. It is these transformed internal coordinates and moment values that are used as input for the NN training. We train a number of nets, each with a different number of hidden nodes, learning rate, and momentum. The net that gives the r^2/q^2 value closest to 1 is used for the moment prediction (see above for refinements). Along with the NN predicted

Table 1. (a) Optimization of the NN Parameters Used To Predict the Monopole Moments of Each Atom in Glycine[a] and (b) NN Architectures Used To Predict the Multiple Moments of C_3

(a)

atom	N_{hidden}[b]	momentum	learning rate	r^2	q^2
C_3	4	0.6	0.05	0.983	0.976
C_4	4	0.7	0.05	0.996	0.995
H_2	9	0.6	0.05	0.999	0.999
H_5	4	0.6	0.05	0.993	0.993
H_6	5	0.9	0.10	0.993	0.992
H_8	8	0.8	0.10	0.999	0.999
H_{10}	6	0.9	0.10	1.000	0.983
N_1	4	0.6	0.05	0.998	0.997
O_7	9	0.7	0.05	0.999	0.999
O_9	5	0.8	0.10	0.998	0.998

(b)

C_3	N_{hidden}[b]	momentum	learning rate	r^2	q^2
q	4	0.6	0.05	0.983	0.976
μ_x	4	0.6	0.05	0.932	0.903
μ_y	4	0.7	0.05	0.940	0.911
μ_z	4	0.6	0.10	0.921	0.880
Q20[c]	4	0.7	0.05	0.978	0.975
Q21c	4	0.6	0.05	0.776	0.627
Q21s	4	0.8	0.05	0.868	0.801
Q22c	4	0.8	0.05	0.986	0.981
Q22s	4	0.6	0.05	0.726	0.627

[a] The correlation coefficients r^2 and q^2 are explained in the main text. [b] Number of nodes in the hidden layer (see eq 7). [c] Notation of ref 29 is adopted for the five quadrupole moments.

moments (monopole, dipole, quadrupole), the energy prediction also uses the higher moments (octupole, hexadecapole) of the optimized (reference) structure. In principle, all moments could have been predicted by NNs, but their energy contributions to 1-n ($n \geq 4$) interactions are small and hence do not justify the concomitant computational overhead.

Nets with 4 to 10 hidden nodes are trained, with momenta of 0.6, 0.7, 0.8, or 0.9, and a learning rate of 0.05 or 0.1. Hence, for each net a total of 7 × 4 × 2 = 56 nets were trained. The number of epochs remained fixed at 100 000 for all training calculations.

4. Results and Discussion

Table 1a shows the parameters of the best NNs used to predict the monopole moments of each atom in glycine. The equivalent data for NMA are not shown since they are very similar. The variation in any of the three parameters has relatively little effect on the monopole prediction quality of the NNs gauged by the r^2 values. The correlation between the monopole predicted by the NNs and the "true" moments calculated by MORPHY is above 0.97 for each atom. However, it is important to note that one cannot predict the best performing NN parameters in advance, just based on the local environment of each atom. The observed optimal number of hidden nodes in the carbon NNs of glycine is 4 or 5 while the range for NMA is 4 to 9. The range for hydrogen is 4 to 10 for both glycine and NMA. The number of nodes for nitrogen in glycine and NMA is, respectively, from 4 to 8 and from 4 to 6. The range for the glycine atoms is 4 to 10 for NMA. The range of parameters used to predict the moments of each atom in glycine and NMA is similar. Any differences are due to the fact that the local chemical environments of the carbon, nitrogen, and oxygen atoms are not exactly alike.

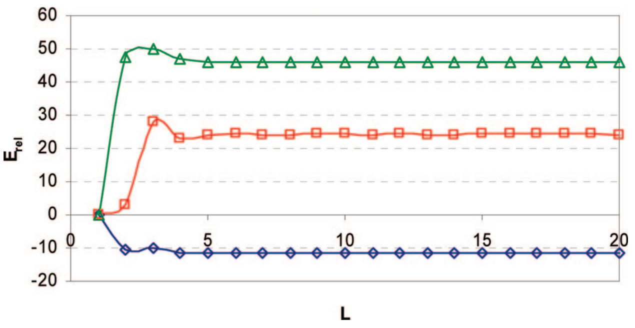


Figure 5. Electrostatic interaction energy (kJ mol^{-1}) of three representative nonbonded interactions in glycine versus rank L; for visual convenience, all energies are relative to the energy at rank $L = 1$ (charge−charge), the largest energy contribution. The N_1-O_9 interaction is marked by red squares, O_7-H_{10} by green triangles, and C_3-H_{10} by blue diamonds.

Table 1b lists the NN parameters used to predict the monopole, dipole, and quadrupole moments of the pivotal α carbon (C_3) in glycine. Each NN has 4 hidden nodes, and 8 out of 9 NNs have a learning rate of 0.05 (the other is 0.1). The monopole moment has the highest r^2 and q^2. Typically, the r^2 or q^2 values decrease as the rank of the moment increases. However, the r^2 and q^2 drops below 0.9 only for some of the quadrupole moments. The Q21c and Q22s moments have the lowest r^2 and q^2 values. Though we do not list the optimal NN parameters of NMA here, their values and the r^2 and q^2 values are similar to those of glycine. In NMA, like in glycine, some quadrupole moments (of C_1) have r^2 and q^2 values below 0.9.

Prediction of multipole moments using NNs and internal coordinates as input is clearly successful, as judged by inspecting the values of the moments themselves. A less direct but physically more valuable way of assessing the performance of the NNs is to look at interatomic interaction energies.[109] The advantage of this assessment is that one can gauge the NN's quality in terms of a single number, namely, energy. The disadvantage is the need to identify other atoms as partners interacting with the central atom for which the NN is set up. Before we analyze the NNs according to errors in energy we must address the technical but paramount issue of convergence control.

A truncated multipole expansion always produces an error compared to the exact interaction energy. Previously we have monitored[110] how energies vary with (expansion) rank L. Factors such as internuclear distance, relative orientation, atomic shape, and the magnitude of the electronic density all determine if the expansion is convergent or not. Comparing the energy profiles calculated from the true multipole moments with the profiles from NN predicted moments reveals how much of the energy differences are due to the prediction error, rather than to an inherent convergence error. To investigate the convergence behavior of the 23 atom−atom interactions in glycine, the energy of each interaction was calculated up to $L = 20$. Figure 5 shows the interaction

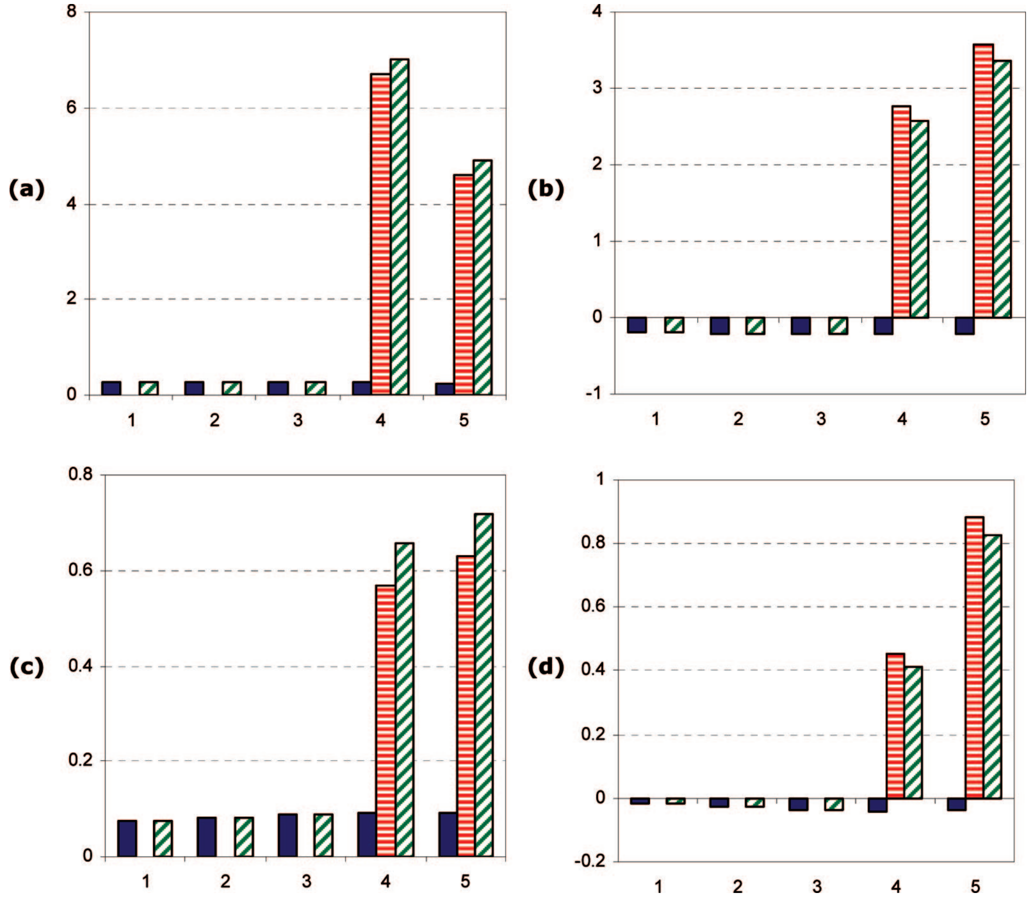


Figure 6. Interaction energy (kJ mol^{-1}) as a function of expansion rank L for four 1-4 atom−atom interactions. (a) N_1−O_9 in glycine, (b) O_7−O_{10} in glycine, (c) O_2−H_8 in NMA, (d) C_3−C_9 in NMA. All energies are relative to the expansion energy calculated using the true moments (monopole, dipole, quadrupole, octupole, and hexadecapole). The blue bar represents the energy calculated from NN-predicted moments (monopole, dipole, and quadrupole) and (fixed) higher moments from the reference geometry. The red bar represents the energy calculated from the *true* monopole, dipole, and quadrupole moments, without higher moments. The green bar is the energy calculated from the NN-predicted moments only (monopole, dipole, and quadrupole), without higher moments.

energy as L is increased for three representative atom−atom interactions. All interactions studied here are convergent including N_1−O_9, the highest energy (501 kJ mol^{-1}) interaction, and O_7−H_{10}, the lowest (−447 kJ mol^{-1}). Of all the interactions studied, the O_7−H_{10} interaction also has the shortest internuclear distance in the reference geometry (2.27 Å).

At rank $L = 1$ the energy represents the charge−charge interaction, $L = 2$ also includes the energy of the charge−dipole interaction, and L=3 includes the charge−quadrupole and dipole−dipole interaction energy. Figure 5 shows how the interaction energy behaves asymptotically beyond $L = 5$, where the change in interaction energy is exceedingly small, less than 0.02 kJ mol^{-1}. This indicates that the interaction has practically converged. Although Figure 5 only shows three energy profiles, these are representative of 1-4 interactions in glycine and NMA. By visual inspection *all* interactions converged in glycine and NMA.

Figure 6 shows the interaction energy as a function of expansion rank L for four 1-4 atom−atom interactions, namely, N_1−O_9 (2.73 Å, termini) in glycine, O_7−H_{10} (2.27 Å, carboxy group) in glycine, and O_2−H_8 (3.11 Å, peptide bond) and C_3−C_9 (3.79 Å, C_α-C'_α) in NMA. The blue bars in the histograms demonstrate how well the NNs predict the monopole, dipole, and quadrupole moments. The bars indicate the energy difference between the true energy (obtained with all true moments up to hexadecupole) and the energy calculated from the NN predicted moments (with true octupole and hexadecupole). All energy differences are of the order of tenths of a kJ mol^{-1} or less. They vary very little with the expansion rank L. This means that already at $L = 1$ (charge−charge) most of the energy difference is present.

Figure 6 also shows the effect on the energy of removing the higher moments (octupole and hexadecupole), as marked by the red bars (horizontal stripes). They show how the energy deviates from the true one, when using the true monopole, dipole, and quadrupole moment. This deviation can be as high as over 6 kJ mol^{-1} for the N_1−O_9 interaction in glycine, which is reaffirmed in the O_2−C_9 interaction in NMA (not shown). Fortunately, the deviation is about an order of magnitude smaller in the C_α−C'_α interaction in NMA, as well as in the peptide bond (O_2−H_8). The green bars are the equivalent of the red ones, where the true monopole, dipole, and quadrupole moments are replaced by the NN predicted ones. The green bars echo what was concluded based on the red ones.

Now that we have shown that the nonbonded interactions are convergent we can use predicted moments in the multipolar expansion to calculate the interaction energies. We assess the performance of the NN by measuring the discrepancy between true and predicted energy. Each of the 23 (nonbonded) interactions in glycine is monitored by an energy error distribution over the external validation set of 50 distorted molecular geometries. The quality of the 50 predictions for each interaction can be measured by the minimum and maximum error, the error range, and the error averaged over the 50 predictions. This information is given in Table 2 for glycine and for the 37 nonbonded interactions in NMA. The glycine interactions cover an energy range of 948 kJ mol^{-1}. The most attractive interaction is O_7-H_{10} (carboxylic oxygen and hydroxyl hydrogen) with an energy of -447 kJ mol^{-1}, while the most repulsive interaction is between N_1 and O_9 (501 kJ mol^{-1}). The N_1-O_9 interaction has the largest average absolute energy difference (E_{diff}) of 2.7 kJ mol^{-1}. However, this represents a difference of just 0.5% between the true energy (E_{true}) and predicted energy ($E_{\mathrm{predicted}}$). Although this interaction has the greatest E_{diff} it is a small fraction of the total interaction energy (E_{int}). The interactions of glycine have an average E_{diff} of 0.2 kJ mol^{-1}, and the largest E_{diff} is just under 1 kJ mol^{-1}. Three interactions have an E_{diff} of 0.5 kJ mol^{-1} or greater, and four have an E_{diff} greater than 0.2 kJ mol^{-1}. A general observation is that as E_{int} increases so does E_{diff}. The range of energy differences ("E_{diff} range" in Table 2) gives an indication of how consistent the energy prediction is. Ideally E_{diff} and the E_{diff} range would both be zero but this is not the case. Again, as E_{diff} increases so does the "E_{diff} range".

The nonbonded interactions of NMA have an energy range of 281 kJ mol^{-1}, which is more than three times smaller than the range for glycine. The average E_{diff} is 0.2 kJ mol^{-1}, which is remarkably similar to glycine's value. The most attractive interaction is O_2-H_8 (located in the peptide bridge), which has an E_{int} of -217 kJ mol^{-1}. The most repulsive interaction is C_1-H_{12} (64 kJ mol^{-1}). Here the interaction with the second highest E_{int} value, O_2-C_9, has the highest E_{diff} value, that is, 1.1 kJ mol^{-1}. The C_3-H_8 interaction has the second highest E_{diff} of 0.5 kJ mol^{-1}. As with glycine NMAs, E_{diff} and "E_{diff} range" increase as the absolute value of E_{int} increases.

Overall, Table 2 shows that the predicted energies are in good agreement with the true energy. Also, the predicted moments can reproduce interactions over a large energy range. For example, the O_7-H_{10} of glycine has an E_{int} of -447 kJ mol^{-1}, but O_7 also interacts with N_1 and this interaction has an energy of 437 kJ mol^{-1}. This corresponds to an energy range of 884 kJ mol^{-1}.

Table 2. Interaction Energy (E_{int}), Absolute Energy Difference between True and Predicted Interaction Energy (E_{diff}), Averaged over a Set of 50 Distorted Geometries, Minimum and Maximum Absolute Energy Difference, and the Range of Interaction Energy Differences of (a) Glycine and (b) *N*-Methylacetamide Molecules[a]

interaction	E_{int}	E_{diff}	Min E_{diff}	max E_{diff}	E_{diff} range
			(a)		
C3−H10	85.92	0.62	0.07	1.86	1.79
C4−H8	219.27	0.35	0.01	1.44	1.42
C4−H2	273.27	0.54	0.00	1.59	1.59
H2−H10	85.94	0.04	0.00	0.36	0.36
H2−H5	14.86	0.05	0.00	0.15	0.14
H2−H6	9.50	0.07	0.00	0.19	0.19
H2−O7	−157.23	0.06	0.00	0.69	0.69
H2−O9	−208.27	0.22	0.02	1.10	1.08
H5−H10	15.77	0.05	0.00	0.20	0.19
H5−H8	15.83	0.08	0.00	0.24	0.24
H5−O7	−45.93	0.16	0.00	0.60	0.60
H5−O9	−38.03	0.13	0.00	0.48	0.48
H6−H10	7.55	0.06	0.00	0.17	0.17
H6−H8	9.26	0.08	0.00	0.19	0.19
H6−O7	−22.62	0.19	0.01	0.48	0.47
H6−O9	−19.97	0.17	0.00	0.53	0.53
H8−H10	64.34	0.03	0.00	0.16	0.16
H8−O9	−147.92	0.10	0.00	0.38	0.38
N1−H10	−212.53	0.17	0.00	0.59	0.59
N1−O7	437.17	0.36	0.03	1.15	1.12
N1−O9	501.26	0.97	0.01	2.70	2.69
O7−H10	−446.56	0.31	0.02	0.57	0.55
O7−H8	−137.97	0.06	0.00	0.23	0.23
average	13.17	0.21	0.01	0.70	0.69
min	−446.56	0.03	0.00	0.15	0.14
max	501.26	0.97	0.07	2.70	2.69
range	947.82	0.94	0.07	2.56	2.55
			(b)		
C1−H10	43.60	0.34	0.01	1.56	1.54
C1−H11	32.65	0.21	0.00	2.40	2.40
C1−H12	63.93	0.34	0.00	1.51	1.50
C3−C9	−7.60	0.26	0.03	0.93	0.90
C3−H10	−1.04	0.03	0.00	0.21	0.21
C3−H11	−0.78	0.03	0.00	0.15	0.15
C3−H12	−1.60	0.04	0.00	0.25	0.25
C3−H8	−15.23	0.53	0.00	1.64	1.63
H4−C9	8.66	0.05	0.00	0.19	0.19
H4−H10	1.14	0.01	0.00	0.06	0.06
H4−H11	0.86	0.01	0.00	0.10	0.10
H4−H12	1.72	0.01	0.00	0.06	0.06
H4−H8	13.37	0.07	0.00	0.23	0.23
H4−N7	−39.48	0.27	0.01	1.09	1.08
H5−C9	3.94	0.06	0.00	0.22	0.22
H5−H10	0.49	0.01	0.00	0.04	0.04
H5−H11	0.39	0.01	0.00	0.07	0.07
H5−H12	0.74	0.01	0.00	0.07	0.07
H5−H8	5.83	0.13	0.00	0.52	0.52
H5−N7	−18.99	0.38	0.03	1.51	1.49
H6−C9	6.98	0.06	0.00	0.22	0.22
H6−H10	0.91	0.01	0.00	0.05	0.05
H6−H11	0.69	0.01	0.00	0.06	0.06
H6−H12	1.36	0.02	0.00	0.10	0.10
H6−H8	11.19	0.11	0.00	0.42	0.42
H6−N7	−32.52	0.34	0.02	1.20	1.18
H8−H10	13.25	0.09	0.00	0.49	0.49
H8−H11	10.84	0.09	0.00	0.98	0.98
H8−H12	17.63	0.07	0.00	0.23	0.22
O2−C9	−201.58	1.12	0.01	3.49	3.49
O2−H10	−29.48	0.24	0.00	1.21	1.20
O2−H11	−22.30	0.13	0.01	1.75	1.74
O2−H12	−45.29	0.27	0.00	0.94	0.93
O2−H4	−53.31	0.30	0.00	1.13	1.13
O2−H5	−24.25	0.28	0.00	0.97	0.96
O2−H6	−40.84	0.43	0.02	1.58	1.56
O2−H8	−217.36	0.25	0.01	0.82	0.82
average	−13.82	0.18	0.00	0.77	0.76
min	−217.36	0.01	0.00	0.04	0.04
max	63.93	1.12	0.03	3.49	3.49
range	281.29	1.12	0.03	3.45	3.45

[a] All energies in kJ mol^{-1}.

Now we look at *the sum* of all nonbonded interaction energies of glycine and NMA for each of the 50 distorted molecules. Figure 7 shows the absolute energy difference (E_{diff}) between the total expansion and the total predicted nonbonded energy as well as the percentage difference. The total nonbonded interaction energy ("E_{true} total") of the 50 distorted glycine molecules has a range of 143 kJ mol^{-1} (from 229 to 371 kJ mol^{-1}). The total predicted nonbonded interaction energy ("$E_{\mathrm{predicted}}$ total") has a range of 143 kJ mol^{-1} (from 227 to 370 kJ mol^{-1}). The average energy difference (E_{diff}) is -0.3 kJ mol^{-1}, and the minimum and maximum E_{diff} values are -4.7 and 4.0 kJ mol^{-1}, respectively. NMA has an average E_{true} of -511 kJ mol^{-1} and a range of 156 kJ mol^{-1} (from -591 to -435 kJ mol^{-1}). The average E_{diff} for each interaction of NMA is 0.2 kJ mol^{-1}. The minimum and maximum E_{diff} values are 0.0 and 3.5 kJ mol^{-1}, respectively.

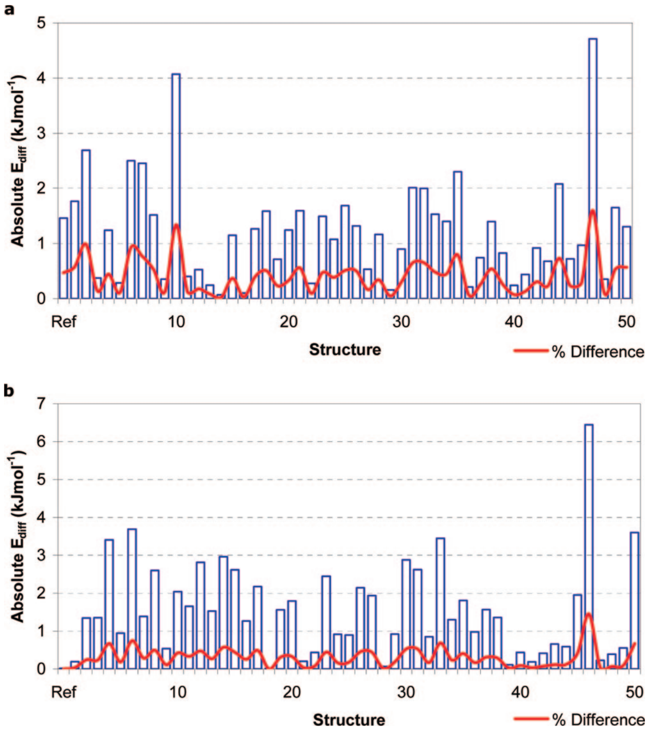


Figure 7. *Total* absolute interaction energy difference of each of the 50 distorted glycine molecules for 1-*n* interactions ($n \geq 4$). The absolute energy difference is $|E_{true} - E_{predicted}|$, where E_{true} is the expansion energy calculated using the true moments and $E_{predicted}$ is the energy calculated using the NN predicted moments. The percentage difference refers to absolute differences.

We initially carried out this work at the HF/6-31G(d) level of theory. The HF method is known to overestimate charge transfer in molecules, witnessed by the exaggerated atomic charges compared to those calculated with Møller−Plesset perturbation theory.[111] We have also observed an increase in QCT charges in going from a B3LYP wave function to a HF one. These increased charges mostly lead to increased interaction energies. For example, the total nonbonded interaction energy for the local reference structure of glycine increases from 310 kJ mol^{-1} using B3LYP moments to 658 kJ mol^{-1} using HF moments, and the energy range increases from 143 kJ mol^{-1} to 250 kJ mol^{-1}. This increase is most noticeable when the individual interactions are considered. The average C_3-H_{10} interaction energy of glycine increases from 86 kJ mol^{-1} to 155 kJ mol^{-1}, a 80% increase, which is the largest change observed. However, not all of the interactions increase in energy. If the interaction has an absolute value less than 50 kJ mol^{-1}, then the interaction increases in magnitude; that is, repulsive interactions become more repulsive and attractive interactions more attractive. However, as these interactions are smaller in magnitude the cumulative effect of using B3LYP moments rather than HF moments is to lower the overall interaction energies for glycine and NMA.

The current method uses NNs to capture the relationship between the local chemical environment of an atom and the polarization of that atom. We now look at how this internal polarization affects the interaction energy. We calculated the energy *including* polarization using the original (i.e., “true”) moments calculated for each of the 50 distorted geometries of glycine. These energies are our reference. To calculate the energy *excluding* polarization we use the moments from the local reference structure in place of the true moments of the distorted molecules. The reference moments are rotated into the correct local frame in each molecule before the energy is calculated. The polarization contribution to the total nonbonded energy and to each of the 23 interactions in glycine is shown in Figure 8.

The polarization energy (E_{pol}) shows great variation. In 12 out of 23 interactions E_{pol} is negative; that is, the energy excluding polarization is higher. In 11 interactions E_{pol} is less than 1 kJ mol^{-1}, and only three interactions (four when including the total interactions) have an E_{pol} greater than 5 kJ mol^{-1}. The E_{pol} for the total nonbonded interaction energy is 12.7 kJ mol^{-1} or 4.2%. The origin of 90% of E_{pol} is due to the difference in charge−charge energy between the two atoms. This is seen for each interaction regardless of the types of atoms that are interacting. Hence, interactions between atoms with large charge differences will have the largest E_{pol}. The N_1-O_9 interaction has an E_{pol} of 17.4 kJ mol^{-1} which is the largest E_{pol} for any interaction (including the total energy). For atoms which have similar charges, such as C and H, E_{pol} is small.

The current method has advantages and limitations. An advantage is that the approach is not limited to a single method for predicting the moments. Alternative machine learning methods can be used in place of the backpropagation NNs, which we have started to investigate. Second, the atoms and their polarization are independent of each other. So, adding new atoms or updating existing ones can be done without having to modify the existing atoms. For example, there is no need to retrain the NNs for a carbonyl carbon when retraining the oxygen because the training sets are independent of each other. Moreover, different moments can be predicted by different machine learning techniques. Third, charge transfer, which normally receives a separate treatment compared to dipolar polarization, is now a unified and streamlined part of general multipolar polarization. A current limitation is the CPU cost of the data preparation and training. The exact amount of time needed depends on the number of inputs required to maintain a ratio of 10 examples for each NN weight. Generating the wave functions accounts for ~25% of the total time needed, calculation of the moments ~65%, and the NN training the remaining ~10%. Another current limitation is the lack of optimization of the polarization procedure. Just to demonstrate the proof-of-concept of this novel technique we use all internal coordinates as input for the moment prediction. In some cases this may mean that each atom is given more information than is necessary to correctly predict that atom’s moments.

In terms of future work, we have just begun to address these limitations. The ultimate goal is to describe polarizable electrostatics in arbitrary amino acids and peptide bonds. Using all internal coordinates, in light of our ultimate goal, means that the NNs are not transferable to other similar atoms in different molecules. Transferability of atom types is a key feature of force fields, and using transferable NNs is one of our aims. Creation of transferable NNs will require a method

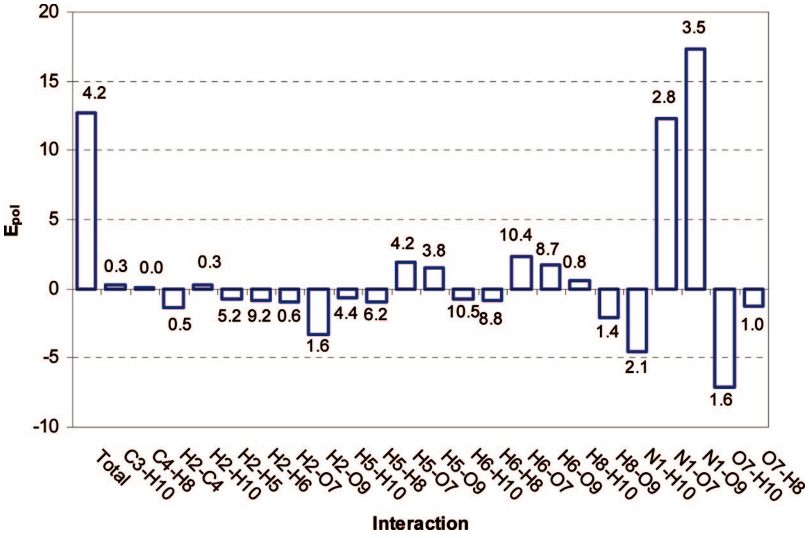


Figure 8. Average energy difference (E_{pol} in kJ mol^{-1}) between the true and the nonpolarized interaction energies of 50 distorted glycine molecules. The numbers above the bars mark the average percentage difference.

for feature selection before the NN training. This is currently being investigated. Also, we have not yet investigated the performance of the predicted moments over the torsional space covered by known protein structures. We are in the process of testing the performance of our method using larger structures. We should also mention that we have obtained successful preliminary results of the current methodology being tested on water clusters. It is important that this methodology will be able to tackle amino acids and water in a unified manner to deliver a reliable protein solvation model.

5. Conclusions

Intramolecular polarization changes an atomic electron density in response to a variation in the positions of the neighboring atoms. We propose a novel method that captures this change directly, by following the change in the atomic multipole moments that express the atomic electron density. Quantum chemical topology provides the partitioning scheme to define the atoms. On the basis of the pilot systems glycine and NMA, we prove that NNs are able to predict with good accuracy the change in atomic monopole, dipole, and quadrupole moments upon changes in molecular configuration. This approach puts charge transfer and dipolar (and high rank multipolar) polarization on a common footing. The current proof-of-concept opens the avenue for a realistic and integrated methodology for a peptide/protein force field, with "informed" atoms that can rapidly and correctly adjust their electronic features to a given environment.

Acknowledgment. We thank the EPSRC for financial support (partially through EP/C015231/1) and GlaxoSmithKline (Stevenage, GB) for additional funding.

Appendix

It is convenient to place the adaptation of the threshold on the same footing as that of the weights. This is simply achieved by thinking of the threshold as an extra weight that is driven by an (artificial) input constantly tied to the value −1. This leads to the negative of the threshold being called the *bias.* The transfer function limits the output of the node to a set range, typically between 0 and 1 or between −1 and +1. To modify the weights of the NN is presented a training data set, with known inputs and outputs. The weights of the NN are modified based upon the error between the prediction made for the examples in the training set and the actual desired output. This type of training is known as *supervised learning*. Training continues for a number epochs as the weights are changed and the NN learns from the training set.

The error based upon the output of the final layer of the NN is fed back through the NN, determining the local gradients of the network as we step back through the network. We have to make use of this gradient as we have no knowledge about the true output for each of the hidden layers. This method is known as *back-propagation of errors.* This process is continued for every example in the data set in an effort to produce results converging upon the function that replicates the relationship between the inputs and the outputs. To begin training the network the weights are randomly initialized. The first step is the network performing a forward pass on a training example and finding a final prediction y. From this the error is calculated via eq A1,

$$e_k(n) = d_k(n) - y_k(n) \tag{A1}$$

where $e_k(n)$ is the output error and $d_k(n)$ is the desired output. This error is then sent back through the network,and the local gradient is calculated, if ρ is set equal to 1 in eq 4,

$$\frac{\partial y_k(n)}{\partial a_k(n)} = \frac{\exp(-a_j(n))}{[1 + \exp(-a_j(n))]^2} \tag{A2}$$

$$\frac{\partial y_k(n)}{\partial v_k(n)} = y_k(n)[1 - y_k(n)] \tag{A3}$$

$$\delta_k^{(L)}(n) = e_k^{(L)}(n)\frac{\partial y_k(n)}{\partial a_k(n)} \tag{A4}$$

The above is only applicable to the final layer of neurons as $e_k^{(L)}$ cannot be defined for a hidden layer. Instead the local gradient is found via eq A5,

$$\delta_j(n) = \frac{\partial y_j(n)}{\partial a_j(n)} \sum_k \delta_k(n)\, w_{kj}(n) \qquad (A5)$$

or it can be expressed as the following equation:

$$\delta_j(n) = y_j(n)[1 - y_j(n)] \sum_k \delta_k(n)\, w_{kj}(n) \qquad (A6)$$

The calculation of the local gradients can then be used to determine the changes to the weights.

$$w_{kj}^{(l)}(n+1) = w_{kj}^{(l)}(n) + \alpha[w_{kj}^{(l)}(n) - w_{kj}^{(l)}(n-1)] + \eta\delta_j^{(l)}(n)\, y_i^{(l-1)}(n) \qquad (A7)$$

where α is the momentum and η is the learning rate. The learning rate controls the magnitude of the changes made to the weights and has a range between 0 and 1. Large changes to the weights mean that the network is not trapped in local minima on the error surface, though the minima may be missed. However, if smaller changes are used then there is less chance that the true minima is missed, though training is slower. The momentum links changes to the weights to the change that took place previously. This ensures that the weights are changed by the necessary magnitude. For instance, if the two previous weights were of the same sign then a large change is expected; however, if the signs were different then the changes made are small. The momentum term accelerates the gradient descent or stabilizes the learning in regions where the sign oscillates.

Standardization is a transformation of the data to dimensionless data, shown in eq A8,

$$x_{\mathrm{st},i} = \frac{(x_i - \bar{x})}{\sigma_n} \qquad (A8)$$

where the standard deviation is $\sigma_n = \sqrt{[(1/N)\sum_{i=1}^{N}(x_i - \bar{x})^2]}$ and $\bar{x}$ is the mean. It is easy to prove that the mean of the standardized data is zero and their standard deviation one.

References

(1) Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A. *J. Phys. Chem. A* **2004**, *108*, 621.

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

(3) Price, S. L.; Harrison, R. J.; Guest, M. F. *J. Comput. Chem.* **1989**, *10*, 552.

(4) Lehn, J. M. *Science* **2002**, *295*, 2400.

(5) Price, S. L. *CrystEngComm* **2004**, *6*, 344.

(6) Sokalski, W. A.; Keller, D. A.; Ornstein, R. L.; Rein, R. *J. Comput. Chem.* **1993**, *14*, 970.

(7) Dobson, C. M. *Nature* **2003**, *426*, 884.

(8) 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.

(9) Jorgensen, W. L. *J. Chem. Theory Comput.* **2007**, *3*, 1877.

(10) Gresh, N.; Kafafi, S. A.; Truchon, J.-F.; Salahub, D. R. *J. Comput. Chem.* **2004**, *25*, 823.

(11) Kaminsky, J.; Jensen, F. *J. Chem. Theory Comput.* **2007**, *3*, 1774.

(12) Rasmussen, T. D.; Ren, P.; Ponder, J. W.; Jensen, F. *Int. J. Quantum Chem.* **2006**, *107*, 1390.

(13) Brooks, B. R.; Brucoleri, R. E.; Olafson, B. D.; Slater, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.

(14) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Profetajr, S.; Wiener, P. *J. Am. Chem. Soc.* **1984**, *106*, 765.

(15) van Gunsteren, W. F.; Berendsen, H. J. C. *GROMOS*; Groningen, The Netherlands, 1987.

(16) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657.

(17) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1984**, *5*, 129.

(18) Cox, S. R.; Williams, D. E. *J. Comput. Chem.* **1981**, *2*, 304.

(19) Besler, B. H.; Merz, K. M.; Kollman, P. A. *J. Comput. Chem.* **1990**, *11*, 431.

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

(21) Chirlian, L. E.; Francl, M. M. *J. Comput. Chem.* **1987**, *8*, 894.

(22) Francl, M. M.; Chirlian, L. E. The pluses and minuses of mapping atomic charges to electrostatic potentials. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; Wiley: New York, 2000; Vol. 14; p 1.

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

(24) Reynolds, C. A.; Essex, J. W.; Richards, W. G. *J. Am. Chem. Soc.* **1992**, *114*, 9075.

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

(26) Mahoney, M. W.; Jorgensen, W. L. *J. Chem. Phys.* **2000**, *112*, 8910.

(27) Nada, H.; van der Eerden, J. P. J. M. *J. Chem. Phys.* **2003**, *118*, 7401.

(28) Vinter, J. G. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 653.

(29) Stone, A. J. *The Theory of Intermolecular Forces*; Clarendon: Oxford, U.K., 1996.

(30) Stone, A. J. *Chem. Phys. Lett.* **1981**, *83*, 233.

(31) Ren, P.; Ponder, J. W. *J. Phys. Chem. B* **2003**, *107*, 5933.

(32) Gordon, M. S.; Freitag, M. A.; Bandyopadhyay, P.; Jensen, J. H.; Kairys, V.; Stevens, W. J. *J. Phys. Chem. A* **2001**, *105*, 293.

(33) Vigne-Maeder, F.; Claverie, P. *J. Chem. Phys.* **1988**, *88*, 4934.

(34) Gresh, N. *J. Comput. Chem.* **1995**, *16*, 856.

(35) Piquemal, J.-P.; Williams-Hubbard, B.; Fey, N.; Deeth, R.; Gresh, N.; Giessner-Prettre, C. *J. Comput. Chem.* **2003**, *24*, 1963.

(36) Piquemal, J.-P.; Cisneros, G. A.; Reinhardt, P.; Gresh, N.; Darden, T. A. *J. Chem. Phys.* **2006**, *124*, 104101.

(37) Gresh, N.; Cisneros, G. A.; Darden, T. A.; Piquemal, J.-P. *J. Comput. Chem.* **2007**, *3*, 1960.

(38) Wannier, G. *Phys. Rev.* **1937**, *52*, 191.

(39) Iannuzzi, M.; Parrinello, M. *Phys. Rev. B* **2002**, *66*, 155209.

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

(41) Sagui, C.; Pomorski, P.; Darden, T. A.; Roland, C. *J. Chem. Phys.* **2004**, *120*, 4530.

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

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

(44) Popelier, P. L. A.; Aicken, F. M. *ChemPhysChem* **2003**, *4*, 824.

(45) Joubert, L.; Popelier, P. L. A. *Mol. Phys.* **2002**, *100*, 3357.

(46) Pilme, J.; Piquemal, J.-P. *J. Comput. Chem.* **2008**, *29*, 1440–1449.

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

(48) Liem, S.; Popelier, P. L. A. *J. Chem. Phys.* **2003**, *119*, 4560.

(49) Liem, S.; Popelier, P. L. A.; Leslie, M. *Int. J. Quantum Chem.* **2004**, *99*, 685.

(50) Liem, S. Y.; Popelier, P. L. A. *J. Chem. Theory Comput.* **2008**, *3*, 353.

(51) Kosov, D. S.; Popelier, P. L. A. *J. Phys. Chem. A* **2000**, *104*, 7339.

(52) Popelier, P. L. A.; Joubert, L.; Kosov, D. S. *J. Phys. Chem. A* **2001**, *105*, 8254.

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

(54) Joubert, L.; Popelier, P. L. A. *Phys. Chem. Chem. Phys.* **2002**, *4*, 4353.

(55) Popelier, P. L. A.; Rafat, M. *Chem. Phys. Lett.* **2003**, *376*, 148.

(56) Rafat, M.; Popelier, P. L. A. *J. Chem. Phys.* **2005**, *123*, 204103.

(57) Rafat, M.; Popelier, P. L. A. *J. Chem. Phys.* **2006**, *124*, 144102.

(58) Rafat, M.; Popelier, P. L. A. Topological atom-atom partitioning of molecular exchange energy and its multipolar convergence. In *Quantum Theory of Atoms in Molecules*; Matta, C. F., Boyd, R. J., Eds.; Wiley-VCH: Weinheim, Germany, 2007; Vol. 5; p 121.

(59) Rafat, M.; Popelier, P. L. A. *J. Comput. Chem.* **2007**, *28*, 832.

(60) Garmer, D. R.; Stevens, W. J. *J. Phys. Chem.* **1989**, *93*, 8263.

(61) Stone, A. J. *Mol. Phys.* **1985**, *56*, 1065.

(62) Angyan, J. G.; Jansen, G.; Loos, M.; Haettig, C.; Hess, B. A. *Chem. Phys. Lett.* **1994**, *219*, 267.

(63) in het Panhuis, M.; Popelier, P. L. A.; Munn, R. W.; Angyan, J. G. *J. Chem. Phys.* **2001**, *114*, 7951.

(64) Houlding, S.; Liem, S. Y.; Popelier, P. L. A. *Int. J. Quantum Chem.* **2007**, *107*, 2817.

(65) Lorenz, S.; Gross, A.; Scheffer, M. *Chem. Phys. Lett.* **2004**, *395*, 210.

(66) Gassner, H.; Probst, M.; Lauenstein, A.; Hermansson, K. *J. Phys. Chem. A* **1998**, *102*, 4596.

(67) Prudente, F. V.; Acioli, P. H.; Soares Neto, J. J. *J. Chem. Phys.* **1998**, *109*, 8801.

(68) Behler, J.; Parrinello, M. *Phys. Rev. Lett.* **2007**, *98*, 146401.

(69) No, K. T.; Chang, B. H.; Kim, S. Y.; Jhon, M. S.; Scheraga, H. A. *Chem. Phys. Lett.* **1997**, *271*, 152.

(70) Cho, K.-H.; No, K. T.; Scheraga, H. A. *J. Mol. Struct.* **2002**, *641*, 77.

(71) Piquemal, J.-P.; Gresh, N.; Giessner-Prettre, C. *J. Phys. Chem. A* **2003**, *107*, 10353.

(72) Spackman, M. A. *Chem. Phys. Lett.* **2006**, *418*, 158.

(73) Friesner, R. A. *Adv. Protein Chem.* **2006**, *72*, 79.

(74) Hodges, M. P.; Stone, A. J. *J. Phys. Chem. A.* **1998**, *102*, 2455.

(75) Yu, H.; van Gunsteren, W. F. *Comput. Phys. Commun.* **2005**, *172*, 69.

(76) Rasmussen, T. D.; Ren, P.; Ponder, J. W. *Int. J. Quantum Chem.* **2007**, *107*, 1390.

(77) Halgren, T. A.; Damm, W. *Curr. Opin. Struct. Biol.* **2001**, *11*, 236.

(78) Batista, E. R.; Xantheas, S. S.; Jónsson, H. *J. Chem. Phys.* **1998**, *109*, 4546.

(79) Coulson, C. A.; Eisenberg, D. *Proc. R. Soc. London, Ser. A* **1966**, *291*, 445.

(80) Handley, C. M.; Popelier, P. L. A. *Synth. React. Inorg. Met.-Org. Nano-Met. Chem.* **2008**, *38*, 91.

(81) Thole, B. T. *Chem. Phys.* **1981**, *59*, 341.

(82) Caldwell, J. W.; Kollman, P. A. *J. Phys. Chem.* **1995**, *99*, 6208.

(83) Gao, J.; Habibollazadeh, D.; Shao, L. *J. Phys. Chem.* **1995**, *99*, 16460.

(84) Soteras, I.; Curutchet, C.; Bidon-Chanal, A.; Dehez, F.; Ángyán, J. G.; Orozco, M.; Chipot, C.; Luque, F. J. *J. Chem. Theory Comput.* **2007**, *3*, 1901.

(85) Cisneros, G. A.; Tholander, S. N.; Parisel, O.; Darden, T. A.; Elking, D.; Perera, L.; Piquemal, J.-P. *Int. J. Quantum Chem.* **2008**, *108*, 1905.

(86) Ledecq, M.; Lebon, F.; Durant, F.; Giessner-Prettre, C.; Marquez, A.; Gresh, N. *J. Phys. Chem. B* **2003**, *107*, 10640.

(87) Chen, W.; Gordon, M. S. *J. Chem. Phys.* **1996**, 105.

(88) Piquemal, J.-P.; Chelli, R.; Procacci, P.; Gresh, N. *J. Phys. Chem.* **2007**, *111*, 8170.

(89) Masia, M. *J. Chem. Phys.* **2008**, *128*, 184107.

(90) Patel, S.; Brooks, C. L. *Mol. Simul.* **2006**, *32*, 231.

(91) Bhat, T. N.; Bentley, G. A.; Boulot, G.; Greene, M. I.; Tello, D.; Dall'Acqua, W.; Souchon, H.; Schwarz, F. P.; Mariuzza, R. A.; Poljak, R. J. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 1089.

(92) Rick, S. W.; Stuart, S. J.; Berne, B. J. *J. Chem. Phys.* **1994**, *101*, 6141.

(93) Harder, E. M. A. V.; Vorobyov, I. V.; Lopes, P. E. M.; Noskov, S. Y.; MacKerell, A. D., Jr.; Roux, B. *J. Chem. Theory Comput.* **2006**, *2*, 1587.

(94) Yu, H.; Hansson, T.; van Gunsteren, W. F. *J. Chem. Phys.* **2003**, *118*, 221.

(95) Yu, H.; van Gunsteren, W. F. *J. Chem. Phys.* **2004**, *121*, 9549.

(96) Yu, H.; Geerke, D. P.; Liu, H.; van Gunsteren, W. F. *J. Comput. Chem.* **2006**, *27*, 1494.

(97) Hemmingsen, L.; Amara, P.; Ansoborlo, E.; Field, M. J. *J. Phys. Chem. A* **2000**, *104*, 4095.

(98) Hagberg, D.; Karlstrom, G.; Roos, B. O.; Gagliardi, L. *J. Am. Chem. Soc.* **2005**, *127*, 14250.

(99) Chen, J.; Martínez, T. J. *Chem. Phys. Lett.* **2007**, *438*, 315.

(100) Gordon, M. S.; Slipchenko, L.; Li, H.; Jensen, J. H. *Annu. Rep. Comput. Chem.* **2007**, *3*, 177.

(101) Gresh, N.; Claverie, P.; Pullman, A. *Int. J. Quantum Chem.* **1982**, *22*, 199.

(102) Popelier, P. L. A. *Mol. Phys.* **1996**, *87*, 1169.

(103) Haettig, C. *Chem. Phys. Lett.* **1996**, *260*, 341.

(104) Vapnik, V. N. *Statistical Learning Theory*; John Wiley: New York, 1998.

(105) Gurney, K. *An Introduction to Neural Networks*; Routledge: London, U.K., 1997.

(106) Haykin, S. *Neural Networks: A Comprehensive Foundation*; Macmillan College Publishing Company: New York, 1994.

(107) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, Revision C.02; Gaussian, Inc.: Wallingford, CT, 2004.

(108) Popelier, P. L. A. *MORPHY98*; Manchester, U.K., 1998.

(109) Rafat, M.; Shaik, M.; Popelier, P. L. A. *J. Phys. Chem. A* **2006**, *110*, 13578.

(110) Rafat, M.; Popelier, P. L. A. *J. Chem. Phys.* **2006**, *124*, 144102.

(111) Velders, G. J. M.; Feil, D. *J. Phys. Chem.* **1992**, *96*, 10725.

CT800166R

JCTC Journal of Chemical Theory and Computation

Geometries of Third-Row Transition-Metal Complexes from Density-Functional Theory

Michael Bühl,*[,†] Christoph Reimann,[‡] Dimitrios A. Pantazis,[‡] Thomas Bredow,[‡] and Frank Neese*[,‡]

School of Chemistry, North Haugh, University of St. Andrews, St. Andrews, Fife KY16 9ST, U.K., and Institut für Physikalische und Theoretische Chemie, Universität Bonn, Wegelerstrasse 12, D-53115 Bonn, Germany

Received May 19, 2008

Abstract: A set of 41 metal−ligand bond distances in 25 third-row transition-metal complexes, for which precise structural data are known in the gas phase, is used to assess optimized and zero-point averaged geometries obtained from DFT computations with various exchange-correlation functionals and basis sets. For a given functional (except LSDA) Stuttgart-type quasi-relativistic effective core potentials and an all-electron scalar relativistic approach (ZORA) tend to produce very similar geometries. In contrast to the lighter congeners, LSDA affords reasonably accurate geometries of 5d-metal complexes, as it is among the functionals with the lowest mean and standard deviations from experiment. For this set the ranking of some other popular density functionals, ordered according to decreasing standard deviation, is BLYP > VSXC > BP86 ≈ BPW91 ≈ TPSS ≈ B3LYP ≈ PBE > TPSSh > B3PW91 ≈ B3P86 ≈ PBE hybrid. In this case hybrid functionals are superior to their nonhybrid variants. In addition, we have reinvestigated the previous test sets for 3d- (Bühl M.; Kabrede, H. *J. Chem. Theory Comput.* **2006**, *2*, 1282−1290) and 4d- (Waller, M. P.; Bühl, M. *J. Comput. Chem.* **2007**, *28*, 1531−1537) transition-metal complexes using all-electron scalar relativistic DFT calculations in addition to the published nonrelativistic and ECP results. For this combined test set comprising first-, second-, and third-row metal complexes, B3P86 and PBE hybrid are indicated to perform best. A remarkably consistent standard deviation of around 2 pm in metal−ligand bond distances is achieved over the entire set of d-block elements.

Introduction

Quantum-chemical calculations require additional approximations to account for relativistic effects when heavier atoms are present. One of the most popular of these approximations is the pseudopotential or effective core potential (ECP) approach,[1] where the innermost electrons are not treated explicitly but subsumed into a specially designed, mean potential acting upon the outer electrons. This ECP can be adjusted numerically such as to account for the leading scalar relativistic effects in the core region even in an otherwise nonrelativistic calculation. Pseudopotentials have fertilized many fields of applied theoretical chemistry and are now in widespread use.

Initially designed at the Hartree−Fock level, ECPs and their corresponding valence basis sets were readily embraced by the ever growing community that uses density functional theory (DFT) in its many flavors. Computational transition-metal chemistry in particular has benefited a lot from this development.[2] From the competing brands of ECPs, two suppliers appear to dominate this market, namely the Hay-Wadt[3] and Stuttgart-Dresden[4] variants,[1] both of which have performed very well in countless validation studies. In contrast, the choice of a suitable exchange-correlation functional from the plethora of vendors is more difficult, first

* Corresponding author fax: +(44)(0)1334 463808; e-mail: buehl@st-andrews.ac.uk (M.B.) and fax: +(49) 228/73-9064; e-mail: neese@thch.uni-bonn.de (F.N.).

[†] University of St. Andrews.

[‡] Universität Bonn.

10.1021/ct800172j CCC: $40.75 © 2008 American Chemical Society
Published on Web 08/22/2008

because of the vast supply of such functionals, and second because their performance may strongly depend on the particular application.

Regardless of their nature, such applications need accurate molecular structures as inputs. We have become interested in assessing the ability of modern DFT methods to reproduce gas-phase geometries of transition-metal complexes in a straightforward, consistent manner. For this purpose, we selected sizable test sets of target molecules, for which reasonably precise and, presumably, accurate structural data are available from gas-phase electron diffraction (GED) or microwave (MW) spectroscopy. In the spirit of Helgaker et al.,[5,6] the performance of several density-functional/basis-set combinations is assessed by correlating computed with experimental bond distances and analyzing the resulting mean and standard deviations. Only bond distances refined experimentally to a precision better than 1 pm are included in this analysis. We have previously reported such assessments for first-[7] and second-row[8] transition-metal complexes, which have revealed subtle differences in the performance of standard density functionals for these two sets. We now extend these studies to complexes from the third transition row. This now allows comprehensive performance tests for computational methods to describe molecular structures that contain metal centers from the whole d-block.

Not only quantum-chemical models such as specific exchange-correlation functionals can be tested this way but also the approximations made to account for relativity. There is growing interest to go beyond the ECP model and to describe all electrons in an explicit relativistic treatment. While full four-component relativistic calculations are still extremely involved and feasible only for atoms and the smallest molecules, two-component variants have evolved to a point that allows their rather routine application to sizable systems. In practice, unless the elements are very heavy the effect of spin−orbit coupling on molecular geometries is limited.[9,10] This suggests that more straightforward and computationally less involved one-component scalar relativistic approaches are the methods of choice for all-electron calculations on third-row transition metals. The advantages of all-electron treatments are obvious if total electron densities are to be computed[11] or−in particular−if spectroscopic properties are computed that depend on the inner-shell electrons or the nodal properties of the valence orbitals. This concerns for example X-ray absorption,[12] Mössbauer[13] and nuclear magnetic[14] or electron paramagnetic resonance[15,16] properties. However, rather special basis sets must be used in all-electron scalar relativistic calculations that are consistent with the relativistic treatment invoked. Such special basis sets have been designed previously for calculations within the Douglas-Kroll-Hess (DKH)[17] or the zeroth order regular approximation (ZORA)[9,18] treatments. However, as far as Gaussian basis sets are concerned, these basis sets are generally contracted and therefore computationally expensive. We have therefore recently reported a series of segmented all electron relativistic (SARC) basis sets for third-row transition metals that can be applied together with the DKH2 and ZORA approaches.[19] Atoms from the first three rows are treated with relativistic recontractions of the Karlsruhe split valence (SV), triple-ζ valence (TZV), or quadruple-ζ valence (QZVP) all-electron basis sets.[20−22]

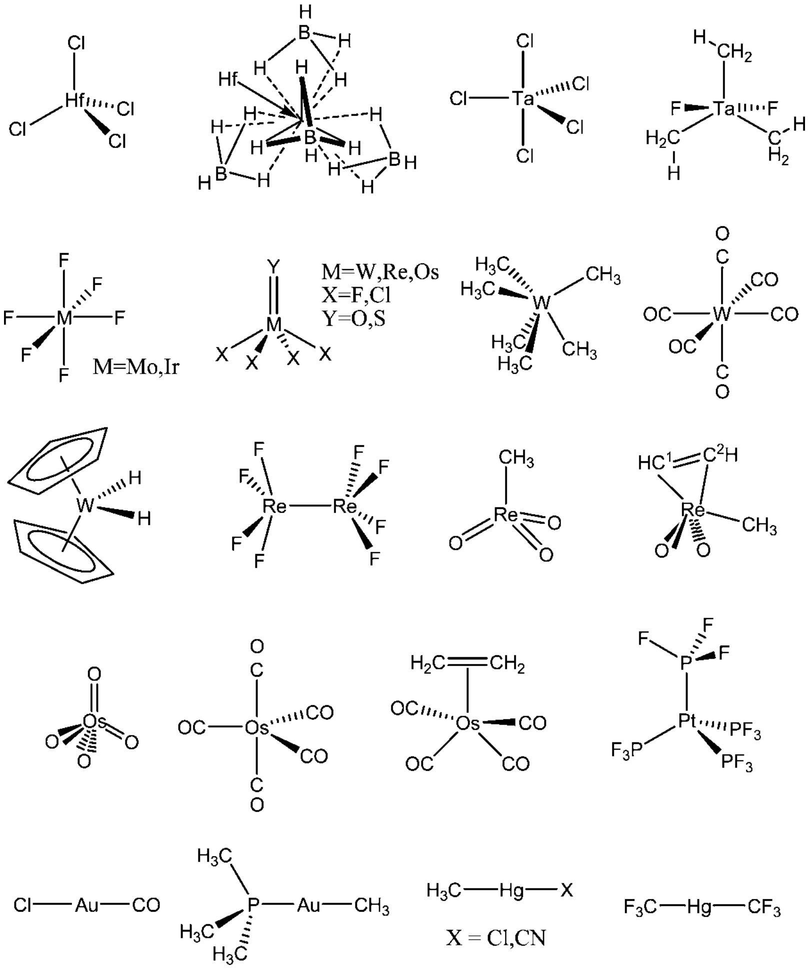
In ref 19, geometries of small transition-metal hydrides, ionization potentials, and binding energies were calculated with the new basis sets and either B3LYP density functional or coupled-cluster with single-, double-, and perturbative triple excitations (CCSD(T)) methods. Here we take the opportunity to compare the performance of all-electron scalar relativistic DFT calculations with ECPs for a much broader range of functionals relative to precise gas-phase structural data of polyatomic molecules. Thus, the present study serves the triple purpose of (a) evaluating the performance of ECP based DFT calculations for the prediction of geometries of third-row transition metals (b) to compare the relative merits of ECP based and scalar-relativistic all-electron calculations throughout the d-block and (c) to evaluate the performance of the SARC all-electron basis sets for 3d-, 4d-, and 5d-transition-metal geometries.

The test set for the 5d metals is shown in Scheme 1. It comprises complexes of the metals from Hf to Hg, for which quite precise experimental data are available from gas-phase electron diffraction (GED) and/or microwave spectroscopy (MW). This test set 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, including complexes with hydride and phosphine ligands and one with a metal−metal bond. Drawing from a large compilation of gas-phase structures,[23] 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 25 molecules with 41 individual bond distances with that precision, which should be sufficient for reasonable statistics. We also report computed zero-point corrections to the bond distances[24,25] for this data set in order to furnish increments to estimate $r_g^{\,0}$ from r_e values,[26] thus facilitating the comparison between theory and experiment.

Computational Details

Geometries were fully optimized in the given symmetry (as given in Table 1) using Gaussian 03[27] and several local (LSDA)[28] and gradient-corrected density functional combinations as implemented therein. Most functionals are composed of one of several exchange parts, namely Becke (B)[29] or Becke hybrid (B3),[30] together with one of several correlation parts, namely Perdew (P86),[31] Perdew−Wang (PW91),[32] or Lee et al. (LYP)[33] (in parentheses: symbols used in combined forms). Other functionals comprise HCTH/407 (denoted HCTH)[34] and the PBE hybrid functional[37] (denoted PBE1, Gaussian keyword PBE1PBE, which is often called PBE0) as well as the meta-GGAs VSXC,[36] TPSS,[37] and TPSS hybrid (denoted TPSSh).[38] 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,[39] and for WMe_6, where spurious imaginary frequencies were found with the default grid; in these cases we used 99 radial shells with 590 angular points. The following relativistic small-core ECPs with the corresponding valence basis sets were employed on the metals: SDD[4] i.e.,

Scheme 1



the Stuttgart-Dresden ECP (together with the [6s5p3d] valence basis) and LANL2DZ[3] (with [3s3p2d] valence basis). On the ligands, the 6−31G∗ basis[40] was used, except for $Hf(BH_4)_4$ and WCp_2H_2, where 6−31G∗∗ was employed for the ligands with a metal−hydrogen bond. In addition, we tested Ahlrichs-type valence basis sets that had been designed for the use with the SDD ECPs,[41] denoted SVP, TZVP, and QZVP (with [5s3p2d1f], [6s4p3d1f], and [7s5p4d3f1g] contractions for the metals, respectively), together with the corresponding all-electron bases on the ligands.[20−22] The minimum character of all optimized structures was verified by evaluation of the harmonic vibrational frequencies at the BP86/SDD level. Closed- and open-shell species were treated with restricted and unrestricted formalisms, respectively. For the computation of effective geometries via the cubic force field, the Barone method[25] was invoked at the BP86/SDD level within Gaussian 03 rev D.01.[27] The default values were used for step size in the numerical differentiation (0.025 Å) and integration grid (SG1).

Scalar relativistic all electron calculations have been performed with the ORCA program package[42] within the ZORA approximation. In our experience ZORA and DKH2 geometries are usually almost indistinguishable. For technical reasons, the relativistic corrections have been performed within the one-center approximation that has previously been shown to be adequate.[43] Geometries have been optimized without constraints due to point symmetry, using the ‘pure’ GGA and meta-GGA functionals (LSDA, BP86, PBE, and TPSS) as well as a variety of hybrid functionals (B3LYP, B3P86, B3PW91, TPSSh, and PBE1). The integration grid was increased to span 80 radial shells and 302 angular grid points. The influence of the empirical van der Waals correction according to Grimme[44] has been studied for BP86, PBE, TPSS, and B3LYP. In all ZORA calculations, the recently published SARC basis sets[19] of TZVP quality has been used for the third-row transition metals and SARC recontractions of the Karlsruhe TZVP basis set for the lighter atoms. For two molecules ($ReOCl_4$ and IrF_6), spin-unrestricted open-shell calculations have been performed.

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. In some cases, not all degrees of freedom have been refined experimentally, or only mean values for formally nonequivalent distances are known to the desired precision. In those cases, we evaluated and assessed the same average of the corresponding optimized parameters, even though full geometry optimizations were performed. This applies to $Os(CO)_5$ and WMe_6. The GED data of the latter were initially refined assuming equal W−C distances; later it was shown that this molecule adopts a structure with lower

Table 1. Bond Lengths *r* (in pm) of Third-Row Transition-Metal Complexes in the Gas Phase[a]

compound (mult.)[b] *sym.*	distance	[bond no.]	reference value	ref	Δr_{vib}
$HfCl_4$ (1) T_d	r(Hf−Cl)	[1]	231.6(5)	46	0.17
$Hf(BH_4)_4$ (1) T	r(Hf−B)	[2]	231.4(2)	47	2.67
	r(Hf−H^{br})	[3]	221.5(7)	47	3.27
$TaCl_5$ (1) D_{3h}	r(Ta−Cl^{mean})	[4]	228.5(2)	48	0.21
$TaMe_3F_2$ (1) C_{3h}	r(Ta−C)	[5]	212.5(5)	49	0.20
	r(Ta−F)	[6]	186.3(4)	49	0.20
WF_6 (1) O_h	r(W−F)	[7]	182.9(2)	50	0.18
WOF_4 (1) C_{4v}	r(W=O)	[8]	166.6(7)	51	0.17
	r(W−F)	[9]	184.7(2)	51	0.22
$WSCl_4$ (1) C_{4v}	r(W=S)	[10]	208.6(6)	52	0.17
	r(W−Cl)	[11]	227.7(3)	52	0.27
WMe_6 (1) C_3	r(W−C)^{mean}	[12]	214.6(3)	53	0.88
$W(CO)_6$ (1) O_h	r(W−C)	[13]	205.9(3)	54	0.40
$W(Cp)_2(H)_2$ (1) C_2	r(W−H)	[14]	170.3(2)	55	0.86
Re_2F_8 (1) D_4	r(Re−Re)	[15]	226.9(5)	56	0.27
	r(Re−F)	[16]	183.0(4)	56	0.20
$ReOCl_4$ (2) C_{4v}	r(Re=O)	[17]	166.3(9)	57	0.10
	r(Re−Cl)	[18]	227.0(5)	57	0.28
ReO_3Me (1) C_{3v}	r(Re=O)	[19]	170.9(3)	58	0.21
	r(Re−C)	[20]	206.0(9)	58	0.41
$ReO_2Me(C_2H_2)$ (1) C_s	r(Re=O)	[21]	171.0(1)	59	0.15
	r(Re−C^{Me})	[22]	211.6(2)	59	0.60
	r(Re−C^1)	[23]	204.3(2)	59	0.54
	r(Re−C^2)	[24]	206.7(2)	59	0.72
OsO_4 (1) T_d	r(Os=O)	[25]	171.2(2)	60	0.27
$OsOCl_4$ (1) C_{4v}	r(Os=O)	[26]	166.3(9)	61	0.11
	r(Os−Cl)	[27]	225.8(5)	61	0.33
$Os(CO)_5$ (1) D_{3h}	r(Os−C)^{mean}	[28]	196.2(4)	62	0.33
$Os(C_2H_4)(CO)_4$ (1) C_{2v}	r(Os−C^{et})	[29]	220.9(5)	63	0.86
	r(Os−C^{ax})	[30]	195.4(2)	63	0.38
	r(Os−C^{eq})	[31]	194.6(5)	63	0.31
IrF_6 (4) O_h	r(Ir−F)	[32]	183.9(2)	50	0.31
$Pt(PF_3)_4$ (1) T_d	r(Pt−P)	[33]	222.9(5)	64	0.53
Au(CO)Cl (1) $C_{\infty v}$	r(Au−Cl)	[34]	221.72(6)	65	0.36
	r(Au−C)	[35]	188.4(2)	65	0.48
$Au(Me)(PMe_3)$ (1) C_3	r(Au−P)	[36]	228.0(5)	66	0.32
Hg(Me)Cl (1) C_{3v}	r(Hg−Cl)	[37]	228.5(3)	67	0.35
	r(Hg−C)	[38]	205.2(5)	67	0.49
$Hg(CF_3)_2$ (1) D_3	r(Hg−C)	[39]	210.6(5)	68	0.40
Hg(Me)(CN) (1) C_{3v}	r(Hg−C^{CN})	[40]	203.69(2)	69	0.39
	r(Hg−C^{Me})	[41]	205.63(1)	69	0.43

[a] Unless otherwise noted, r_a or r_α values from GED are given. [b] (In parentheses: multiplicity) ax = axial, br = bridging, Cp = cyclopentadienyl, eq = equatorial, et = ethylene.

symmetry and two sets of nonequivalent W−C bonds.[53] Allowing for fluxional behavior in the gas phase, the refined mean value is probably sufficiently precise. Another such case is $TaCl_5$, where two GED studies[48a,b] have reported almost identical mean values of equatorial and axial bonds but disagree markedly on their difference (which varies between 4.7 pm[48a] and 14.2 pm[48b]). It is probably the fluxional behavior of this molecule with its very low Berry pseudorotation barrier[48b] that makes the actual precision of the individual bond distances somewhat lower than suggested by the quoted standard deviations (which are all well below our target value). Thus, we only discuss the mean Ta−Cl distance in this case, as this appears to be refined reasonably well and in a reproducible manner. $Pt(PF_3)_4$ is also indicated to be fluxional, since the GED data have been found to be consistent with free rotation about the Pt−P bond.[64] Both staggered and eclipsed conformations turned out to be minima at the BP86(SDD) level, with marginal differences in the optimized bond distances. We employed the slightly more stable eclipsed form[45] throughout this study.

The final selected experimental parameters are collected in Table 1. Most distances are r_a or r_α values determined from GED, and some are r_z or r_0 geometries known from MW spectroscopy. In general, when both sets of parameters are known, they tend to be in very good mutual accord, with differences rarely exceeding 1 pm, our target precision.

Performance of the ECP Models. Individual distances optimized with the various density-functional/ECP/basis-set combinations are given as Supporting Information. The resulting statistical assessment, that is, the mean and standard deviations from the reference data in Table 1, is summarized in Table 2 ($\bar{D}^{equil}$ and $\bar{D}^{equil}_{std}$. values, respectively). Deviations are defined as $r_{calc} - r_{exp}$, such that positive mean deviations denote overestimation of the bond lengths by DFT. In addition, the mean absolute and the maximum errors to either side are included in Table 2 (labeled $|\bar{D}|^{equil}$ and D^{equil}_{max}, respectively). It turned out that, in particular, the standard deviation is strongly influenced by a single outlier, namely the Hf−H^{br} bond in $Hf(BH_4)_4$ (bond no. 3), which is significantly underestimated at all DFT levels.[70] In order to assess the effect of this bond on the overall statistics, we also provide $\bar{D}^{equil}_{std}$ values where this bond has been removed from the data set (values in parentheses in Table 2).

First, all functionals were tested with the SDD ECP and valence basis on the metal and 6−31G* basis on the ligands (entries 1−12 in Table 2, arranged in the order of increasing mean deviation). Next, another ECP and/or other basis sets were employed for selected functionals, notably BP86 (for

Table 2. Statistical Assessment of Equilibrium (r_e) and Effective (r_{eff}) Metal−Ligand Bond Distances Computed for the Test Set in Scheme *1* at a Number of Levels of Theory[a]

entry	functional	ECP/basis set[b]	$\bar{D}^{equil}$	$\lvert\bar{D}^{equil}\rvert$	$\bar{D}^{equil}_{std}$ [c]	D^{equil}_{max}	$\bar{D}^{eff}$	$\bar{D}^{eff}_{std}$
1	LSDA	SDD	−0.26	1.56	2.10 (1.46)	−8.7 [3]	0.23	1.70
2	PBE1	SDD	1.07	1.67	2.07 (1.58)	5.8 [36]	1.56	1.76
3	B3P86	SDD	1.32	1.83	2.09 (1.53)	−7.7 [3]	1.81	1.75
4	B3PW91	SDD	1.64	2.05	2.11 (1.60)	−7.1 [3]	2.13	1.80
5	TPSSh	SDD	2.24	2.66	2.25 (1.53)	−8.2 [3]	2.72	1.88
6	PBE	SDD	2.76	3.06	2.17 (1.69)	6.9 [36]	3.25	1.82
7	B3LYP	SDD	2.92	3.22	2.43 (1.97)	9.6 [36]	3.41	2.18
8	TPSS	SDD	2.94	3.33	2.33 (1.59)	−7.8 [3]	3.43	1.94
9	BPW91	SDD	3.05	3.34	2.24 (1.74)	7.6 [36]	3.54	1.90
10	BP86	SDD	3.10	3.39	2.21 (1.69)	7.5 [36]	3.59	1.87
11	VSXC	SDD	3.23	3.56	2.51 (1.95)	9.4 [36]	3.72	2.22
12	BLYP	SDD	4.78	5.01	2.63 (2.19)	11.6 [36]	5.27	2.37
13	BP86	LANL2DZ	3.94	4.50	6.09 (5.87)	21.0 [39]	4.43	5.94
14	BP86	LANL2DZ[d]	5.82	6.33	6.56 (6.39)	21.3 [37]	6.31	6.40
15	BP86	SDD/SVP[e]	2.78	3.06	2.17(1.71)	8.0 [36]	3.27	1.90
16	BP86	SDD/TZVP[e]	2.33	2.67	2.21 (1.68)	−6.7 [3]	2.82	1.91
17	BP86	SDD/QZVP[e]	1.89	2.22	1.94 (1.37)	−6.8 [3]	2.37	1.63
18	B3P86	SDD/SVP[e]	0.97	1.85	2.23 (1.79)	−7.5 [3]	1.46	2.00
19	B3P86	SDD/TZVP[e]	0.49	1.56	2.15 (1.63)	−8.4 [3]	0.97	1.88
20	PBE1/	SDD/SVP[e]	0.69	1.74	2.23 (1.85)	−7.3 [3]	1.18	2.02
21	PBE1	SDD/TZVP[e]	0.18	1.48	2.12 (1.66)	−8.2 [3]	0.67	1.87
22	PBE1	SDD/QZVP[e]	−0.23	1.26	1.90(1.42)	−8.3 [3]	0.26	1.66
23	LSDA	SDD/QZVP	−1.49	1.77	1.93 (1.45)	−9.9 [3]	−0.99	1.56
24	LSDA	ZORA/TZVP	−1.82	2.01	2.30 (1.66)	−11.9 [3]	−1.34	1.87
25	PBE1	ZORA/TZVP	−0.50	1.53	2.21 (1.66)	−9.7 [3]	−0.02	1.90
26	B3P86	ZORA/TZVP	0.04	1.60	2.28(1.69)	−9.6 [3]	0.53	1.96
27	B3PW91	ZORA/TZVP	0.07	1.63	2.27 (1.71)	−9.4 [3]	0.56	1.97
28	TPSSh	ZORA/TZVP	0.56	1.72	2.45 (1.67)	−10.7 [3]	1.05	2.08
29	PBE	ZORA/TZVP	1.26	2.08	2.32 (1.74)	−8.5 [3]	1.75	1.97
30	B3LYP	ZORA/TZVP	1.50	2.29	2.60 (2.09)	−8.3 [3]	1.99	2.34
31	TPSS	ZORA/TZVP	1.29	2.08	2.52 (1.70)	−10.4 [3]	1.78	2.13
32	BPW91	ZORA/TZVP	1.52	2.22	2.38 (1.80)	−8.4 [3]	2.01	2.04
33	BP86	ZORA/TZVP	1.50	2.20	2.39 (1.78)	−8.6 [3]	1.99	2.04
34	BLYP	ZORA/TZVP	3.45	3.79	2.79 (2.27)	9.1 [36]	3.93	2.52
35	PBE+VdW	ZORA/TZVP	1.14	2.12	2.49 (1.82)	−9.6 [3]	1.63	2.09
36	B3LYP+VdW	ZORA/TZVP	1.34	2.31	2.78 (2.15)	−9.9 [3]	1.83	2.47
37	TPSS+VdW	ZORA/TZVP	1.12	2.10	2.74 (1.80)	−11.9 [3]	1.61	2.31
38	BP86+VdW	ZORA/TZVP	1.32	2.23	2.61 (1.87)	−10.2 [3]	1.81	2.20
39	BLYP+VdW	ZORA/TZVP	3.25	3.77	3.03 (2.38)	8.8 [36]	3.73	2.69

[a] All values are in picometers relative to experimentally reported values (r_{exp}). $\bar{D}^{equil}$, $\lvert\bar{D}_{equil}\rvert$, $\bar{D}^{equil}_{std}$, and D^{equil}_{max} denote mean, mean absolute, standard, and maximum deviations, respectively, for the equilibrium geometries, $\bar{D}^{eff}$ and $\bar{D}^{eff}_{std}$ are the corresponding deviations for the zero-point averaged, effective geometries. In square brackets: bond numbers from Table 1 for which the maximum error occurs. [b] 6−31G* basis for the ligands, except where otherwise noted. [c] In parentheses: standard deviations for geometries excluding bond no. 3 (see text). [d] D95 for the ligands. [e] The corresponding Ahlrichs basis sets are used on the ligands.

comparison with the results for the first and second transition rows), B3P86, and PBE1.

Following the procedure of our previous studies, effective geometries were then computed at the BP86/SDD level, via numerical computation of the cubic force field using the method of Barone et al. This affords incremental corrections to the bond distances, Δr_{vib} (given in the last column of Table 1), leading from the equilibrium values r_e to the zero-point averaged ones, r_g^0. Arguably, the latter are better suited for direct comparison to the experimental, thermally averaged distances than the former. Actually, there is evidence for small first-row molecules that the zero-point motion affords the largest correction to equilibrium distances and that thermal effects on top of them (i.e., the difference between zero and finite T) tend to be much smaller.[71] If this holds also for the transition-metal complexes, the effective or r_g^0 geometries should be a quite good approximation to the experimental r_a or r_0 structures.

Assuming the same extent of transferability between computational levels that has been established in our studies of 3d- and 4d-metal complexes, we have added the Δr_{vib} values evaluated at the BP86/SDD level to the corresponding equilibrium distances obtained at all other levels and repeated the statistical analysis with respect to the experimental reference data. The corresponding mean and absolute deviations are included in the last two columns in Table 2, labeled $\bar{D}^{eff}$ and $\bar{D}^{eff}_{std}$. The former, mean error is shifted with respect to that of the equilibrium distances, $\bar{D}^{equil}$, by a constant amount of ca. +0.5 pm. This is because all individual increments (last column in Table 1) are positive, i.e. bonds get longer upon zero-point averaging. The individual increments themselves are quite variable, however, ranging from very small changes for metal−oxo multiple bonds (ca. 0.1 pm), via intermediate values for metal−carbon bonds (up to ca. 0.9 pm), to quite large values for the bonds involving the boranate ligand in Hf(BH$_4$)$_4$, where the corrections amount to more than 3 pm for the Hf−H distance (see Table 1). Since this distance appears to be significantly underestimated in most equilibrium geometries (see the Supporting

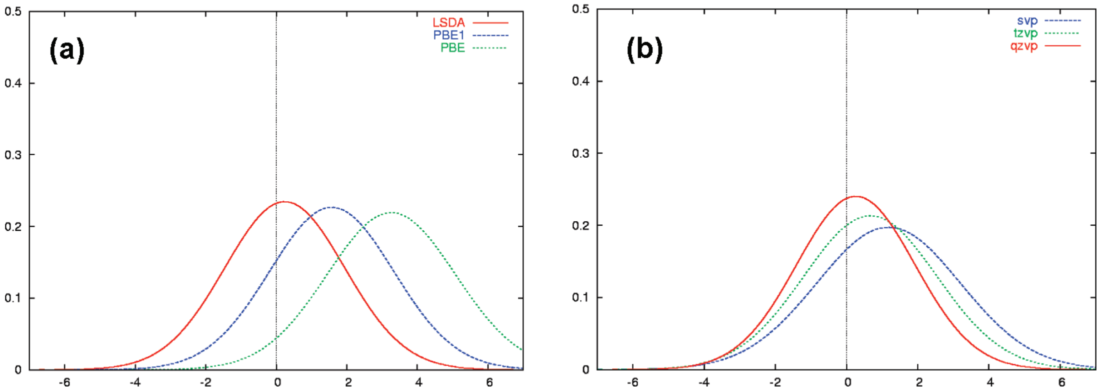


Figure 1. Normal distributions for the errors in the effective bond distances for the test set in Scheme 1. The distributions have been calculated from the mean and standard deviations in Table 2 and are all normalized to one. (a) Left: dependence on the density functional using SDD ECP and valence basis (6−31G∗ on the ligands). (b) Right: dependence on the basis set for the PBE1 hybrid functional together with the SDD ECP.

Information and D_{max}^{equil} values in Table 2),[70] the vibrational correction significantly reduces the error for this bond, thereby leading to noticeable improvements in the standard deviations (compare $\bar{D}_{std}^{equil}$ and $\bar{D}_{std}^{eff}$ values in Table 2).

The following conclusions can be drawn from our results for the 5d-metal complexes:

1. In conjunction with SDD and 6−31G∗ basis, LSDA outperforms all other functionals. It has the smallest mean deviation close to zero for both equilibrium and effective geometries and one of the smallest standard deviations (entry 1 in Table 2). This observation is in marked contrast to the first- and second-row transition-metal complexes, where the tendency of LSDA to overbind translates into optimized (or effective) distances that are much too short.

2. Hybrid functionals are consistently superior to GGAs and meta-GGAs, except for B3LYP, which is surpassed by PBE, and more or less matched by a number of other standard GGAs such as BPW91 or BP86. The two most promising hybrid functionals are PBE1 and B3P86.

3. BLYP and the meta-GGA VSXC produce some of the largest mean and standard deviations and cannot be recommended, consistent with our findings for the lighter metal complexes.

4. The LANL2DZ ECPs together with their compact valence bases are inferior to the corresponding SDD variants with their more flexible basis sets. The large errors evident from Table 2 for LANL2DZ (entry 13) are to a large extent due to some spectacular failures for the linear Hg(II) species in the set (see the Supporting Information and D_{max}^{equil} values in Table 2).[72] For the other complexes, the relative performance of LANL2DZ and SDD is less disparate, but the latter is, in general, slightly superior (arguably due to the more flexible valence basis on the metal).[73]

5. Larger basis sets are beneficial. In particular in the Ahlrichs series, the systematic increase of the metal-valence and ligand bases from SVP to TZVP and QZVP is concomitant with a decrease in mean and standard deviations (e.g., with the PBE1 functional, entries 20−22 in Table 2). For LSDA, such a basis-set extension worsens the agreement with experiment somewhat (compare entries 1 and 23 in Table 2), but also at the LSDA/SDD/QZVP level, a very respectable mean error (below 1 pm for $\bar{D}^{eff}$) and one of the lowest standard deviations remain.

The good performance of LSDA for the 5d complexes is noteworthy. The tendency to underestimate metal−ligand bond lengths at that level is most pronounced in the first transition row,[7] somewhat alleviated but still noticeable in the second,[8] and all but disappeared in the third. This trend seen in the whole sets is also found in individual homologous compounds that are present in all sets, namely the group-4 tetrachlorides and group-8 pentacarbonyls (see selected data in Table S4 in the Supporting Information). For main-group compounds, the ubiquitous overbinding of LSDA does not appear to result in such a pronounced underestimation of bond lengths as found for the 3d-metals (see ref 74 and some illustrative data in Table S5 of the Supporting Information).

To conclude this section, LSDA and most hybrid functionals are quite robust in reproducing geometries of third-row transition-metal complexes and tend to be more accurate than pure or meta-GGAs. Except for LSDA, PBE1 affords the lowest mean deviation and one of the lowest standard deviations, 1.6 and 1.8 pm, respectively, at the SDD level (which are further improved with the larger TZVP and QZVP basis sets). The best GGA is PBE, 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 2 (analogous to the presentation by Helgaker et al.).[5,6] Figure 1b illustrates the basis-set dependence for one particular density functional, PBE1, where increase of the basis results in noticeable shifts of the normal distribution and some reduction of its width for the largest basis, qzvp. The provisional ranking of the functionals for the 5d-metal complexes, ordered according to increasing mean deviation at the SDD level, is thus the following:

LSDA < PBE1 ≈ B3P86 ≈ B3PW91 < TPSSh < PBE ≈ B3LYP ≈ TPSS ≈ BPW91 ≈ BP86 < VSXC < BLYP

Performance of the All-Electron Models. The scalar relativistic results generally show slightly smaller mean deviations than their ECP counterparts (except for LSDA), but the corresponding standard deviations always slightly exceed those of the SDD ECP results. However, for all

Table 3. Statistical Assessment of Equilibrium (r_e) and Effective (r_{eff})[76] Metal−Ligand Bond Distances Computed for the Combined Test Sets of All 3d-, 4d-, and 5d-Metal Complexes at Selected Levels of Theory[a]

entry	functional	3d ECP/basis set[b]	4d,5d ECP/basis set[b]	$\bar{D}^{equil}$	$\|\bar{D}^{equil}\|$	$\bar{D}^{equil}_{std}$	$D^{equil\,c}_{max}$	$\bar{D}^{eff}$	$\bar{D}^{eff}_{std}$
1	BP86	SDD	SDD	1.40	2.41	2.63	7.5 [5d:36]	1.94	2.56
2	BP86	AE1	SDD	1.80	2.37	2.32	7.5[5d:36]	2.34	2.22
3	B3P86	AE1	SDD	0.04	1.60	2.10	−7.7 [5d:3]	0.57	2.02
4	BLYP	AE1	SDD	3.71	3.84	2.61	11.6 [5d:36]	4.25	2.58
5	B3LYP	AE1	SDD	1.85	2.41	2.35	9.6 [5d:36]	2.39	2.35
6	B3LYP	SDD	SDD	1.43	2.45	2.68	9.6 [5d:36]	1.97	2.69
7	BPW91	AE1	SDD	1.78	2.35	2.31	7.6 [5d:36]	2.32	2.20
8	B3PW91	AE1	SDD	0.39	1.67	2.12	−7.1 [5d:3]	0.93	2.05
9	TPSS	AE1	SDD	1.59	2.16	2.27	−7.8 [5d:3]	2.12	2.11
10	TPSSh	AE1	SDD	0.91	1.80	2.18	−8.2 [5d:3]	1.44	2.05
11	LSDA	AE1	SDD	−2.01	2.72	2.71	−8.7 [5d:3]	−1.47	2.54
12	VSXC	AE1	SDD	2.56	2.79	2.48	16.9 [4d:28]	3.10	2.48
13	PBE1	AE1	SDD	−0.17	1.65	2.14	−7.4 [5d:3]	0.37	2.08
14	BP86	SVP	SDD/SVP[d]	1.16	2.13	2.40	8.0 [5d:36]	1.70	2.36
15	BP86	TZVP	SDD/TZVP[d]	1.39	2.04	2.19	−6.7 [5d:3]	1.92	2.12
16	BP86	QZVP	SDD/QZVP[d]	0.93	1.72	1.99	−6.8 [5d:3]	1.47	1.93
17	BP86	TZVP	ZORA/TZVP	1.24	2.05	2.29	−8.6 [5d:3]	1.78	2.18
18	TPSS	TZVP	ZORA/TZVP	1.04	1.83	2.19	−10.4 [5d:3]	1.58	2.02
19	TPSSh	TZVP	ZORA/TZVP	0.31	1.54	2.07	−10.7 [5d:3]	0.84	1.92
20	PBE	TZVP	ZORA/TZVP	0.99	1.93	2.25	−8.5 [5d:3]	1.53	2.13
21	PBE1	TZVP	ZORA/TZVP	−0.79	1.97	2.00	−9.7 [5d:3]	−0.25	1.93
22	PBE+VdW	TZVP	ZORA/TZVP	0.87	1.63	2.40	−9.6 [5d:3]	1.41	2.24
23	LSDA	ZORA/TZVP	ZORA/TZVP	−2.63	2.96	2.65	−11.9 [5d:3]	−2.09	2.46
24	PBE1	ZORA/TZVP	ZORA/TZVP	−1.05	1.81	2.11	−9.7 [5d:3]	−0.51	2.04
25	B3P86	ZORA/TZVP	ZORA/TZVP	−0.48	1.71	2.19	−9.6 [5d:3]	0.06	2.12
26	B3PW91	ZORA/TZVP	ZORA/TZVP	−0.46	1.71	2.18	−9.4 [5d:3]	0.08	2.11
27	TPSSh	ZORA/TZVP	ZORA/TZVP	0.04	1.69	2.23	−10.7 [5d:3]	0.58	2.09
28	PBE	ZORA/TZVP	ZORA/TZVP	0.73	2.00	2.43	−8.5 [5d:3]	1.26	2.32
29	B3LYP	ZORA/TZVP	ZORA/TZVP	1.17	2.15	2.50	−8.3 [5d:3]	1.70	2.50
30	TPSS	ZORA/TZVP	ZORA/TZVP	0.77	1.91	2.38	−10.4 [5d:3]	1.30	2.22
31	BPW91	ZORA/TZVP	ZORA/TZVP	1.00	2.11	2.46	−8.4 [5d:3]	1.54	2.36
32	BP86	ZORA/TZVP	ZORA/TZVP	0.98	2.11	2.47	−8.6 [5d:3]	1.52	2.38
33	BLYP	ZORA/TZVP	ZORA/TZVP	3.11	3.44	2.86	9.1[5d:36]	3.65	2.83
34	PBE+VdW	ZORA/TZVP	ZORA/TZVP	0.62	2.04	2.56	−9.6 [5d:3]	1.16	2.41
35	B3LYP+VdW	ZORA/TZVP	ZORA/TZVP	0.92	2.03	2.48	−9.9 [5d:3]	1.45	2.40
36	TPSS+VdW	ZORA/TZVP	ZORA/TZVP	0.54	1.94	2.53	−11.9 [5d:3]	1.08	2.31
37	BP86+VdW	ZORA/TZVP	ZORA/TZVP	0.83	2.12	2.61	−10.2 [5d:3]	1.36	2.47
38	BLYP+VdW	ZORA/TZVP	ZORA/TZVP	2.78	3.21	2.83	8.8 [4d:28]	3.32	2.73

[a] See footnotes in Table 2. [b] See footnotes in Table 2. [c] In brackets: transition row and corresponding running bond number from refs 7 and 8 and this work. [d] See footnotes in Table 2.

intents and purposes, the results are very similar since the difference in the standard deviations is merely 0.17 pm on average for all methods. The standard deviation is significantly reduced when the Hf−H^{br} bond distance (bond no. 3) is discarded as an outlier (see Table 2, values in parentheses, and the discussion above). The errors for the zero-point averaged effective geometries follow the same trend as described above for the ECP case.

The ranking of the functionals is slightly changed in the all-electron calculations since the LSDA functional now shows one of the largest mean deviations thus indicating a systematic underestimation of bond distances. However, its standard deviation is still quite small. Consistent with the ECP results, the hybrid functionals B3P86, B3PW91, PBE1, and TPSSh provide the most accurate results, while the performance of B3LYP and BLYP is considerably worse. In fact, B3LYP and BLYP exhibit the largest standard deviations in this set. The GGA and meta-GGA functionals are found to give similar results, with PBE again being superior to TPSS, BP86, and BPW91. The inclusion of the empirical van der Waals (VdW) corrections does not lead to noticeable improvements in the results in this test set (compare for instance, entries 29 and 35 in Table 2). We have, however, frequently found significantly improved geometries in sterically crowded systems and stacked pi-systems with this correction. Upon adding the zero-point average, the standard deviations are further improved, again in agreement with the ECP results.

Performance of the Models for All Transition Rows. Combining the present results on the third-row metals with those from our previous studies on first- and second-row metals affords a comprehensive validation for the whole d-block. A selection of levels that are available for all sets[75] are assessed in Table 3. For the first and second transition row, additional scalar relativistic ZORA calculations have been performed according to the approach described above, in order to allow for a fair comparison.

Unexpectedly, the standard deviations are somewhat smaller in the all-electron calculations when the 3d-complexes are calculated without relativistic corrections, while the mean errors are superior only for PBE (with and without van der Waals contributions) and PBE1. The effect of the relativistic corrections is to decrease the metal−ligand bond distances. According to our experience the nonrelativistic all-electron calculated DFT distances are slightly overestimated in many Werner type complexes.[77] Hence the scalar relativistic effects will often provide a correction in the right direction. However, for the present set of 3d transition-metal complexes this seems not to be the case. By comparing the AE1(3d)/SDD(4d+5d) to TZVP(3d)/ZORA+TZVP(4d+5d) results in Table 3 for the functionals BP86, TPSS, TPSSh, and PBE1, the errors are slightly reduced, with the mean deviation of PBE1 being the only exception. For BP86, more combinations of methods and basis sets have been evaluated than for the other functionals. Using ECPs for 3d, 4d, and 5d molecules give the largest errors, while a scalar relativistic treatment throughout all 3 rows gives the lowest error but still a rather large standard deviation. The combination AE1(3d)/SDD+TZVP(4d+5d) gives the best standard deviation, while the mean error is slightly larger than the one for the combination TZVP(3d)/ZORA+TZVP(4d+5d).

Because most functionals show subtle differences in performance for the various transition rows (e.g., TPSS is very good for 3d complexes, but lags behind for the heavier congeners), the overall performance of the functionals tends to even out over the whole d-block. BLYP and VSXC show large mean and standard deviations throughout, and for LSDA, the good performance in the ECP calculations for the 5d row cannot make up for the deficiencies apparent for the 3d and 4d series. Overall the latter three functionals are trailing behind the others and cannot be recommended for geometry optimizations of transition-metal complexes. Most of the other functionals form a sort of peloton, for which it is difficult to single out clear leaders. The slight superiority of B3P86 and PBE1 noted in the 4d and 5d complexes is preserved for the whole set, however. Thus, these functionals emerge as being quite robust for the computation of geometries of transition-metal complexes in general.

However, while these functionals do show low mean deviations from experiment (between ca. 0.4 pm and 0.6 pm, $\bar{D}^{eff}$ values in Table 3) and have the lowest associated standard deviations of ca. 2 pm (see $\bar{D}^{equil}_{std}$ or $\bar{D}^{eff}_{std}$ values in Table 3), the latter values imply a notable scatter of the computed bond distances about the experimental values. For comparison, the accuracy achievable with highly sophisticated *ab initio* methods for equilibrium bond distances of light main-group compounds is much better (cf. mean and standard deviation around 0.2 and 0.3 pm, respectively, at CCSD(T)/cc-pVQZ).[5,6] In this context it should be kept in mind that even reasonably precise GED results for transition-metal complexes, which form a major source of the experimental database used in our analyses, 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. Thus, the high accuracy achievable for light main-group compounds appears to be out of reach, or at least undetectable, for transition-metal complexes. Nevertheless, there appears to be room for improvement in the development of new exchange-correlation functionals for the description of transition-metal complexes.

Conclusions

This work concludes our extended validation study of DFT methods for the prediction of transition-metal complex geometries. Together with the data obtained for 3d and 4d transition-metal species[7,8] a rather comprehensive set of data has been assembled that documents the strengths and weaknesses of modern DFT methods for the prediction of transition-metal geometries. It turns out that no single functional is clearly superior to all others, and, hence, a variety of choices remains possible. Overall, there is a slight advantage of hybrid functionals, especially PBE1 (sometimes also called PBE0), and B3P86 or B3PW91 appear to be the most advantageous choices. Since PBE1 has also been found to perform exceedingly well for many other properties including energetics,[78] excitation energies,[79] or EPR properties,[80] it may even be preferred over B3LYP for general chemistry applications. Nevertheless, very significant computational advantages can be realized if nonhybrid (GGA or meta-GGA) functionals are combined with the density fitting technique (a factor of 5−10 represents a typical speedup over conventional implementations). In this respect, the excellent behavior of the PBE functional should be mentioned as a viable alternative. However, it is clearly necessary to proceed to basis sets of at least triple-ζ quality if accurate results are to be obtained. Small, unpolarized basis sets such as LANL2DZ[3] cannot be recommended if it is desired that the results reflect the properties of the functional more than the shortcomings of the basis set used. The extended study also demonstrates that well designed ECPs, such as the Stuttgart/Dresden ones,[4] can safely be used for studying transition-metal complex geometries. All-electron calculations are now equally feasible since suitable segmented Gaussian basis sets of various double-through quadruple-ζ quality are available.[19−22] Their performance in conjunction with the ZORA or DKH2 scalar relativistic treatments is very similar to that in the ECP case without an undue increase in computation time. The exception are hybrid DFT calculations on 5d species where the significant number of f-primitives required to describe the 4f-shell properly does add noticeably to the computational effort. No such bottlenecks arise in nonhybrid calculations within the density fitting approximation, in particular if the efficient Split-RI-J variant is used that behaves particularly well with respect to higher angular momentum basis functions.[81] The advantages of the all-electron treatment become significant upon calculating molecular properties such as total electron-densities,[11] Mössbauer spectra,[13] X-ray absorption spectra,[12] NMR,[14] or EPR spectra.[15,16]

Acknowledgment. M.B. wishes to thank EaStChem for support and H. Früchtl for technical assistance. Computations were performed on a Fujitsu Siemens PC and an Opteron PC cluster at the University of St. Andrews. We acknowledge financial support from the University of Bonn as well as the SFB 624, the SFB 663, and the priority program 1137. We also thank Mr. Jens Mekelburger for technical assistance.

Supporting Information Available: Bond distances of the 5d test set in Table 1 and of the 3d and 4d test sets, optimized at selected levels, and PBE1/QZVP optimized geometries of the 5d set. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) See for instance: Dolg, M. In *Modern Methods and Algorithms of Quantum Chemistry, Proceedings,* 2nd ed.; Grotendorst, J., Ed.; John von Neumann Institute for Computing: Jülich, Germany, 2000; NIC Series Vol. 3, pp 507−540; www.fz-juelich.de/nic-series/Volume3/dolg.pdf and the extensive bibliography cited therein.

(2) E.g., Frenking, G.; Antes, I.; Böhme, M.; Dapprich, S.; Ehlers, A. W.; Jonas, V.; Neuhaus, A.; Otto, M.; Stegmann, R.; Veldkamp, A.; Vyboishchikov, S. F. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH Publishers: New York, 1996; Vol. 8, pp 63−144.

(3) Hay, P. J.; Wadt, W. R. *J. Chem. Phys.* **1985**, *82*, 299–310.

(4) Dolg, M.; Wedig, U.; Stoll, H.; Preuss, H *J. Chem. Phys.* **1987**, *86*, 866–872.

(5) Helgaker, T.; Gauss, J.; Jørgensen, P.; Olsen, J. *J. Chem. Phys.* **1997**, *106*, 6430–6440.

(6) Bak, K. L.; Gauss, J.; Jørgensen, P.; Olsen, J.; Helgaker, T.; Stanton, J. F. *J. Chem. Phys.* **2001**, *114*, 6548–6556.

(7) (a) Bühl, M.; Kabrede, H. *J. Chem. Theory Comput.* **2006**, *2*, 1282–1290. (b) Waller, M. P.; Bühl, M. *J. Comput. Chem.* **2007**, *28*, 1531–1537.

(8) Waller, M. P.; Braun, H.; Hojdis, N.; Bühl, M. *J. Chem. Theory Comput.* **2007**, *3*, 2234–2242.

(9) van Lenthe, E.; Snijders, J. G.; Baerends, E. J. *J. Chem. Phys.* **1996**, *105*, 6505–6516.

(10) van Wüllen, C.; Langermann, N. *J. Chem. Phys.* **2007**, *126*, 114106.

(11) Vyboishchikov, S. F.; Frenking, G. *J. Comput. Chem.* **1997**, *18*, 416–429.

(12) (a) DeBeer-George, S.; Petrenko, T.; Neese, F. *Inorg. Chim. Acta* **2008**, *361*, 965–972. (b) DeBeer-George, S., Petrenko, T.; Neese, F. submitted to *J. Phys. Chem.* (c) Ray, K.; Petrenko, T.; Wieghardt, K. *Dalton Trans.* **2007**, 1552–1566. (d) Kokatam, S.; Ray, K.; Pap, J.; Bill, E.; Geiger, W. E.; LeSuer, R. J.; Rieger, P. H.; Weyhermüller, T.; Neese, F.; Wieghardt, K. *Inorg. Chem.* **2007**, *46*, 1100–1111. (e) Ray, K.; DeBeer-George, S.; Solomon, E. I.; Wieghardt, K.; Neese, F. *Chem. Eur. J.* **2007**, *13*, 2783–2797. (f) Kapre, R.; Ray, K.; Sylvestre, I.; Weyhermüller, T.; DeBeer-George, S.; Neese, F.; Wieghardt, K. *Inorg. Chem.* **2006**, *45*, 3499–3509.

(13) (a) Neese, F. *Inorg. Chim. Acta* **2002**, *337C*, 181–192. (b) Sinnecker, S.; Slep, L.; Bill, E.; Neese, F. *Inorg. Chem.* **2005**, *44*, 2245–2254. (c) Ray, K.; Begum, A; Weyhermüller, T.; Piligkos, S.; van Slageren, J.; Neese, F.; Wieghardt, K. *J. Am. Chem. Soc.* **2005**, *127*, 4403–4415. (d) Ray, K.; Weyhermüller, T.; Neese, F.; Wieghardt, K. *Inorg. Chem.* **2005**, *44*, 5345–5360.

(14) For some recent reviews see e.g.: (a) Bühl, M. *Annu. Rep. NMR Spectrosc.* In press. (b) Autschbach, J. *Coord. Chem. Rev.* **2007**, *251*, 1796–1821. (c) Autschbach, J. *Struct. Bonding (Berlin)* **2004**, *112*, 1–48.

(15) (a) Fritscher, J.; Hrobarik, P.; Kaupp, M. *J. Phys. Chem. B* **2007**, *111*, 4616–4629. (b) Munzarova, M.; Kaupp, M. *J. Phys. Chem. A* **1999**, *103*, 9966–9983.

(16) (a) Neese, F. *J. Chem. Phys.* **2001**, *115*, 11080–11096. (b) Neese, F. *J. Chem. Phys.* **2003**, *117*, 3939–3948. (c) Neese, F. *J. Chem. Phys.* **2007**, *127*, 164112. (d) Neese, F. *J. Am. Chem. Soc.* **2006**, *128*, 10213–10222. (e) Sun, X.; Chun, H.; Hildenbrand, K.; Bothe, E.; Weyhermüller, T.; Neese, F.; Wieghardt, K. *Inorg. Chem.* **2002**, *41*, 4295–4303.

(17) Hess, B. A.; Marian, C. M. In *Computational Molecular Spectroscopy*; Jensen, P., Bunker, P. R., Eds.; John Wiley & Sons: New York, 2000; p 169ff.

(18) (a) Heully, J. L.; Lindgren, I.; Lindroth, E.; Martenssonpendrill, A. M. *Phys. Rev. A* **1986**, *33*, 4426–4429. (b) van Lenthe, J. G.; Baerends, E. J.; Snijders, E. *J. Chem. Phys.* **1993**, *99*, 4597–4610. (c) van Wüllen, C. *J. Chem. Phys.* **1998**, *109*, 392–399.

(19) Pantazis, D. A.; Chen, X.-Y.; Landis, C. R.; Neese, F. *J. Chem. Theory Comput.* **2008**, *4*, 908–919.

(20) Schäfer, A.; Horn, H.; Ahlrichs, R. *J. Chem. Phys.* **1992**, *97*, 2571–2577.

(21) Schäfer, A.; Huber, C.; Ahlrichs, R. *J. Chem. Phys.* **1994**, *100*, 5829–5835.

(22) Weigend, F.; Furche, F.; Ahlrichs, R. *J. Chem. Phys.* **2003**, *119*, 12753–12762.

(23) *Landolt-Börnstein, Structure Data of Free Polyatomic Molecules*; Kuchitsu, K., Ed.; Springer Verlag: Berlin, 1998; New Series, Vol. II/25.

(24) (a) Ruud, K.; Åstrand, P.-O; Taylor, P. R. *J. Chem. Phys.* **2000**, *112*, 2668–2683. (b) Ruud, K.; Åstrand, P.-O; Taylor, P. R. *J. Am. Chem. Soc.* **2000**, *123*, 4826–4833. (c) Ruden, T.; Lutnæss, O. B.; Helgaker, T. *J. Chem. Phys.* **2003**, *118*, 9572–9581.

(25) (a) Barone, V. *J. Chem. Phys.* **2004**, *120*, 3059–3065. (b) Barone, V. *J. Chem. Phys.* **2005**, *122*, 014108.

(26) The equilibrium distance, r_e, is the distance between the positions of the nuclei on the potential energy surface, as obtained from standard geometry optimizations; r_g is the average internuclear distance at temperature T; and r_g^0 is the average internuclear distance at 0 K. It is the latter value that our computed effective geometries refer to. Typical quantities derived experimentally are r_a (the effective internuclear distance as derived from electron scattering intensity), r_α (the distance between average nuclear positions in the thermal equilibrium at temperature T), r_z (the distance between average nuclear positions in the ground vibrational state), or r_0 (the effective internuclear distance obtained from the rotational constants), 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.

(27) *Gaussian 03, Revision D.01*; Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, R.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.;

Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. Gaussian, Inc.: Wallingford, CT, 2004.

(28) Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200–1211. Functional III of that paper used.

(29) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098–3100.

(30) Becke, A. D. *J. Chem. Phys.* **1996**, *98*, 5648–5642.

(31) (a) Perdew, J. P. *Phys. Rev. B* **1986**, *33*, 8822–8824. (b) Perdew, J. P. *Phys. Rev. B* **1986**, *34*, 7406.

(32) Perdew, J. P. In *Electronic Strucure of Solids*; Ziesche, P., Eischrig, H., Eds.; Akademie Verlag: Berlin, 1991. (b) Perdew, J. P.; Wang, Y. *Phys. Rev. B* **1992**, *45*, 13244–13249.

(33) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789.

(34) Boese, A. D.; Handy, N. C. *J. Chem. Phys.* **2001**, *114*, 5497–5503.

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

(36) Van Voorhis, T.; Scuseria, G. E. *J. Chem. Phys.* **1998**, *109*, 400–410.

(37) (a) Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. *Phys. Rev. Lett.* **2003**, *91*, 146401. (b) Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. *Phys. Rev. Lett.* **2004**, *120*, 6898–6911.

(38) (a) Staroverov, V. N.; Scuseria, G. E.; Tao, J.; Perdew, J. P. *J. Chem. Phys.* **2003**, *119*, 146401. (b) Staroverov, V. N.; Scuseria, G. E.; Tao, J.; Perdew, J. P. *J. Chem. Phys.* **2004**, *121*, 11507.

(39) Johnson, E. R.; Wolkow, R. A.; DiLabio, G. A. *Chem. Phys. Lett.* **2004**, *394*, 334–338.

(40) (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.

(41) Weigend, F.; Ahlrichs, R. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305.

(42) Neese, F. *ORCA - an ab initio, Density Functional and Semiempirical Program Package, 2.6−35*; Universität Bonn: Bonn, Germany, 2008.

(43) van Lenthe, E.; Faas, S.; Snijders, J. G. *Chem. Phys. Lett.* **2000**, *328*, 107–112.

(44) (a) Grimme, S. *J. Comput. Chem.* **2004**, *25*, 1463–1476. (b) Grimme, S. J. *Comput. Chem.* **2006**, *27*, 1787–1799.

(45) E.g., the eclipsed form is more stable than the staggered one by 3.7 and 4.3 kcal/mol at BP86/SDD and B3LYP/SDD levels, respectively.

(46) Girichev, G. V.; Petrov, V. M.; Giricheva, N. I.; Utkin, A. N.; Petrova, V. N. *Russ. J. Struct. Chem.* **1981**, *22*, 694.

(47) Borisenko, K. B.; Downs, A. J.; Robertsen, H. E.; Rankin, D. W. H.; Tang, C. Y. *Dalton Trans.* **2004**, 967–970.

(48) (a) Faegri Jr, K.; Haaland, A.; Martinsen, K.-G.; Strand, T. G.; Volden, H. V.; Swang, O.; Anderson, C.; Persson, C.; Bogdanovic, S.; Herrmann, W. A. *J. Chem. Soc., Dalton Trans.* **1997**, *101*, 3–1018. For an earlier study see:(b) Ischenko, A. A.; Strand, T. G.; Demidov, A. V.; Spiridonov, V. P. *J. Mol. Struct.* **1978**, *43*, 227.

(49) Kadel, J.; Oberhammer, H. *Inorg. Chem.* **1994**, *33*, 3197–3198.

(50) Richardson, A. D.; Hedberg, K.; Lucier, G. M. *Inorg. Chem.* **2000**, *39*, 2787–2793.

(51) Robiette, A. G.; Hedberg, K.; Hedberg, L. *J. Mol. Struct.* **1977**, *37*, 105–112.

(52) Page, E. M.; Rice, D. A.; Hagen, K.; Hedberg, L.; Hedberg, K. *Inorg. Chem.* **1982**, *21*, 3280.

(53) (a) Original GED study assuming D_{3h} symmetry: Haaland, A.; Hammel, A.; Rypdal, K.; Volden, H. V. *J. Am. Chem. Soc.* **1990**, *112*, 4547–4549. (b) An irregular prism with two nonequivalent sets of W−C distances differing by ca. 6−8 pm has been found in the solid state: Kleinhenz, S.; Pfennig, V.; Seppelt, K. *Chem. Eur. J.* **1998**, *4*, 1687–1691. (c) This irregular prism has been found by means of DFT computations: Kaupp, M. *Chem. Eur. J.* **1998**, *4*, 1678–1686.

(54) Arnesen, S. P.; Seip, H. M. *Acta Chem. Scand.* **1966**, *20*, 2711.

(55) MW, r_0 value: Tackett, B. S.; Karunatilaka, C.; Daly, A. M.; Kukolich, S. G. *Organometallics* **2007**, *26*, 2070–2076.

(56) Giricheva, N. I.; Girichev, G. V.; Lapshina, S. B.; Shl'ykov, S. A.; Politov, Yu. A.; Butskii, V. D.; Pervov, V. S. *Russ. J. Struct. Chem. (Engl. Transl.)* **1993**, *34*, 214–224.

(57) Hagen, K.; Hobson, R. J.; Rice, D. A.; Turp, N. *J. Mol. Struct.* **1985**, *128*, 33–40.

(58) (a) GED values from the following: Herrmann, W. A.; Kiprof, P.; Rypdal, K.; Tremmel, J.; Blom, R.; Alberto, R.; Behm, J.; Albach, R. W.; Bock, H.; Solouki, B.; Mink, J.; Lichtenberger, D.; Gruhn, N. E. *J. Am. Chem. Soc.* **1991**, *113*, 6527–6537. (b) MW results ("best fit" values for Re=O and Re−C distances of 170.3(2) and 207.4(4) pm, respectively) see: Wikrent, P.; Drouin, B. J.; Kukolich, S. G.; Lilly, J. C.; Ashby, M. T.; Herrmann, W. A. *J. Chem. Phys.* **1997**, *107*, 2187–2192.

(59) MW, r_0 structure: Kukolich, S. G.; Drouin, B. J.; Indris, O.; Dannemiller, J. J.; Zoller, J. P.; Herrmann, W. A. *J. Chem. Phys.* **2000**, *113*, 7891–7900.

(60) Seip, H. M.; Stølevik, R. *Acta Chem. Scand.* **1966**, *20*, 385.

(61) Hagen, K.; Hobson, R. J.; Holwill, C. J.; Rice, D. A. *Inorg. Chem.* **1986**, *25*, 3659–3661.

(62) Huang, J.; Hedberg, K.; Pomeroy, R. K. *Organometallics* **1988**, *7*, 2049–2053.

(63) MW, r_0 structure: Karunatilaka, C.; Tackett, B. S.; Washington, J.; Kukolich, S. G. *J. Am. Chem. Soc.* **2007**, *129*, 10522–10530.

(64) Ritz, C. L.; Bartell, L. S. J. *Mol. Struct.* **1976**, *31*, 73–76.

(65) MW, r_0 structure: Evans, C. J.; Reynard, L. M.; Gerry, M. C. L. *Inorg. Chem.* **2001**, *40*, 6123–6131.

(66) Haaland, A.; Hougen, J.; Volden, H. V.; Puddephatt, R. J. *J. Organomet. Chem.* **1987**, *325*, 311–315.

(67) MW, r_s structure: Walls, C.; Lister, D. G.; Sheridan, J. *J. Chem. Soc., Faraday Trans. II* **1975**, *71*, 1091–1099.

(68) Oberhammer, H. *J. Mol. Struct.* **1978**, *48*, 389–394.

(69) MW: Cox, A. P.; Rego, C. A. *J. Chem. Phys.* **1988**, *89*, 124–128.

(70) A notable underestimation of this bond length has also been noted at the MP2/LANL/6−3+G∗ level, from which some geometrical and force-field parameters have been used during

refinement of the GED data, cf. ref 47. In that paper an unusually high vibrational amplitude has been noted for this bond and has been attributed to a fluxional process exchanging bridging and terminal H atoms. If the potential energy surface for this process were highly anharmonic with a very low barrier, a theoretical description of the light H atoms might require quantum dynamical methods; however, we could find no evidence for this, because a notable barrier of 6.9 and 5.9 kcal/mol is indicated at the BP86/SDD and B3LYP/SDD levels, respectively (including zero-point energies), proceeding via a Hf(η^3-BH_4)$_3$(η^2-BH_4) transition state.

(71) Toyama, M.; Oka, T.; Morino, Y. *J. Mol. Spectrosc.* **1964**, *13*, 193–213.

(72) Only minor improvements are brought about by the use of an f-function on the metal taken from the following:Hollwart, A.; Böhme, M.; Dapprich, S.; Ehlers, A. W.; Gobbi, A.; Jonas, V.; Kohler, K. F.; Stegmann, R.; Veldkamp, A.; Frenking, G. *Chem. Phys. Lett.* **1993**, *208*, 237–240. *Chem. Phys. Lett.* **1994**, *224*, 603.

(73) Very recently, and after this work was started, more flexible basis sets have been devised for the Hay-Wadt ECPs, the combination thereof denoted LANL2TZ: Roy, L. E.; Hay, P. J.; Martin, R. L. *J. Chem. Theory Comput.* **2008**, *4*, 1029–1031.

(74) See: Koch, W.; Holthausen, M. C. *A Chemist's Guide to Density Functional Theory*, 2nd ed.; Wiley-VCH: Weinheim, 2001; and the extensive bibliography therein.

(75) Because we had not tested the PBE1 functional, which performs so well for the geometries of the heavier metal complexes, in our initial study on the 3d congeners, we have now reoptimized the latter set at the PBE1/AE1 level. For this set alone, this level affords mean and standard deviations of−1.4 and 1.8 pm, respectively, for the equilibrium geometries, and−0.9 and 1.8 pm, respectively, for the effective geometries.

(76) The rovibrational corrections for $Co(CO)_3(NO)$ were erroneously given as zero in Table 1 of ref 7b, whereas they should read 0.45 and 0.64 pm for the Co−N and Co−C bonds, respectively. We apologize for this oversight, which affects the final assessment of the whole set only marginally, and does not alter any of the qualitative conclusions.

(77) (a) Neese, F. *J. Biol. Inorg. Chem.* **2006**, *11*, 702–711. (b) Neese, F. *Coord. Chem. Rev.* **2008**, in press, DOI: 10.1016/j.ccr.2008.05.014.

(78) Grimme, S. *J. Phys. Chem. A* **2005**, *109*, 3067–3077.

(79) (a) Adamo, C.; Scuseria, G. E.; Barone, V. *J. Chem. Phys.* **1999**, *111*, 2889–2899. (b) Adamo, C.; Barone, V. *Theor. Chem. Acc.* **2000**, *105*, 169–172.

(80) (a) Neese, F. *J. Chem. Phys.* **2001**, *115*, 11080. (b) Kossmann, S.; Kirchner, B.; Neese, F. *Mol. Phys.* **2007**, *105*, 2049–2071.

(81) Neese, F. *J. Comput. Chem.* **2003**, *24*, 1740–1747.

CT800172J

JCTC Journal of Chemical Theory and Computation

Torsional Barriers and Equilibrium Angle of Biphenyl: Reconciling Theory with Experiment

Mikael P. Johansson*,†,‡ and Jeppe Olsen†

Lundbeck Foundation Center for Theoretical Chemistry, Aarhus University, DK-8000 Århus C, Denmark, and Department of Chemistry, University of Helsinki, Helsinki, Finland

Received May 22, 2008

Abstract: The barriers of internal rotation of the two phenyl groups in biphenyl are investigated using a combination of coupled cluster and density functional theory. The experimental barriers are for the first time accurately reproduced; our best estimates of the barriers are 8.0 and 8.3 kJ/mol around the planar and perpendicular conformations, respectively. The use of flexible basis sets of at least augmented quadruple-ζ quality is shown to be a crucial prerequisite. Further, to finally reconcile theory with experiment, extrapolations of both the basis set toward the basis set limit and electron correlation toward the full configuration-interaction limit are necessary. The minimum of the torsional angle is significantly increased by free energy corrections, which are needed to reach an agreement with experiment. The density functional B3LYP approach is found to perform well compared with the highest level *ab initio* results.

1. Introduction

From an electronic structure point of view, biphenyl ($C_{12}H_{10}$, see Figure 1) is a surprisingly challenging molecule. Especially pinpointing the energetics of the internal rotation around the central C−C bond connecting the two benzene units has proven problematic. The traditional picture of the interactions involved in deciding the torsional angle between the two twisted phenyl planes is that of an energetic competition between the favorable π-conjugation between the two planes and the steric repulsion between the adjacent hydrogens in *ortho*-position. Here we should mention that this interpretation of the counterbalancing interactions was recently challenged by Matta et al.,[1] who proposed that the hydrogen−hydrogen interaction would in fact be attractive and that the reason for nonplanarity instead is caused by an unfavorable lengthening of the central C−C bond when the planes become more coplanar. This view was later rejected in favor of the classical interpretation by Poater et al.[2] The debate is ongoing.[3] Anyhow, opposing interactions are involved, and therefore a theoretical approach that treats all important effects on equal footing is necessary for a reliable description of the potential energy surface.

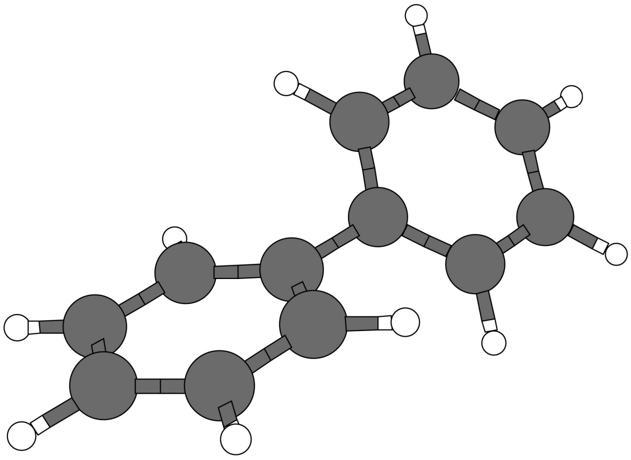


Figure 1. The equilibrium structure of biphenyl, $C_{12}H_{10}$. The figure was created with XMakemol.[4]

In the most recent gas-phase experiments, Bastiansen and Samdal estimated the barriers to be 6.0 ± 2.1 kJ/mol and 6.5 ± 2.0 kJ/mol around 0° and 90°, respectively,[5] while Almenningen et al. found the equilibrium angle to be 44.4 ± 1.2°.[6] The computational reproduction of the experimental barriers of torsion has hitherto proved to be difficult for

* Corresponding author e-mail: mikael.johansson@iki.fi.

† Aarhus University.

‡ University of Helsinki.

10.1021/ct800182e CCC: $40.75 © 2008 American Chemical Society
Published on Web 08/08/2008

theoretical methods. *Ab initio* quantum chemical methods tend to give too high barriers, especially for the rotation around the planar conformation.[7−15] Recently, Sancho-García and Cornil[16] performed a thorough and systematic study of the energetics of the torsional potential of biphenyl. Using high-level correlated wave function methods, their best estimates of the barriers were still higher than those deduced from experiment. The 0° barrier was 2.4 kJ/mol beyond the experimental uncertainty. More importantly, the order of the barrier heights differed from experiment, with ΔE (0°) > ΔE (90°). The elusiveness of the experimental values has led to speculations about possible problems and ambiguities in the experimental interpretation.

While experimental problems naturally cannot be ruled out and were commented on already in the original work,[5] we set our goal to obtain a theoretical treatment that is as thorough as possible, by performing high-level electronic structure calculations, including all major effects. Hence, we explore the limits of all three dimensions of quantum chemical accuracy: *(i)* the amount of correlation energy accounted for, *(ii)* the completeness of the one-particle basis set (which Sancho-García and Cornil[16] identified as probably the largest source of error left in their study), and, last and least, *(iii)* the completeness of the Hamiltonian, that is, inclusion of relativity. As we will show, meticulous calculations allow us to finally reconcile theory with experiment, for the right reasons, without relying on compensating errors.

2. Methodology Overview

The geometries of biphenyl in various conformations have been obtained at the density functional theory (DFT) level,[17,18] within the generalized gradient approximation (GGA),[19] using the popular combination of Becke's three-parameter hybrid exchange functional[20] in connection with the Lee−Yang−Parr correlation[21] functional, B3LYP; the correlation of the uniform electron gas was modeled with the Vosko-Wilk-Nusair VWN5 formulation.[22] The doubly polarized triple-ζ quality basis-set, TZVPP,[23] was used during optimization. Final B3LYP and Hartree−Fock (HF)[24,25] energies were evaluated from extrapolated energies using Jensen's polarization consistent basis set series, pc-*n*.[26−29] Scanning of the potential energy surface (PES) of the relative torsional angle of the phenyl planes was done by optimizing all coordinates except the torsional angle. Additional geometry optimizations were performed using HF and second order Møller−Plesset perturbation theory (MP2),[30] within the density-fitting resolution of the identity formulation (RI-MP2).[31,32]

Correlated *ab initio* wave function (WF) energies were calculated using the B3LYP structures at the following levels of theory: MP2, the spin component scaled version of MP2 (SCS-MP2),[33] and coupled cluster (CC) including single and double excitations, CCSD,[34] as well as perturbative triples corrections, CCSD(T).[35] In general, the 1*s* orbitals of the carbon atoms were kept frozen, and Dunning's standard basis sets[36,37] of up to the augmented quadruple-ζ level, i.e., 1420 basis functions, were employed. Fully correlated calculations were performed with the weighted core-valence basis set of Peterson and Dunning.[38] Details on the use of basis sets are given in the discussion. The basis set limit was estimated by the two-point scheme of Halkier et al.[39] (eq 3). The full configuration-interaction limit was extrapolated from the CC values with Goodson's continued fraction method[40] (eq 4).

The zero point energies (ZPE) as well as the enthalpies ΔH and free energies ΔG at experimental temperature were estimated within the harmonic approximation, treating rotation and translation classically. The vibrational frequencies were calculated analytically[41] at the B3LYP/TZVPP level. Relativistic effects were computed at the B3LYP level with the one-step exact two-component relativistic Hamiltonian recently presented by Iliaš and Saue.[42]

Molecular symmetry was exploited to speed up the calculations. The planar (0°) conformation was assigned D_{2h} symmetry, and the perpendicular (90°) conformation D_{2d} (abelian C_{2v} where necessary). The intermediate torsion angle conformer calculations were performed in D_2 symmetry.

The correlated wave function calculations were performed with the Molpro 2006.1 package;[43−45] the Turbomole 5.91 program suite[46−51] was used for nonrelativistic DFT and all geometry optimizations; and relativistic calculations were performed with the Dirac package.[52] Default convergence and threshold parameters were employed, with the following, tighter exceptions: The Molpro aug-cc-pVQZ calculations used one- and two-electron integral thresholds of 10^{-15}; the Turbomole calculations used the "m4" type grid[53] and a self-consistent field (SCF) convergence criterion of 10^{-7} Hartree. Some basis sets were obtained via the convenient Basis Set Exchange portal.[54,55]

3. Results and Discussion

In this section, we begin by examining the pure, nonrelativistic electronic energies at 0 K of biphenyl. The relative energies of three conformations are studied: the equilibrium structure as well as the planar and perpendicular transition states. After that, various correction terms to the energies are discussed. These include relativistic effects, zero-point vibrational energies and thermal corrections, and the effect of correlating all electrons. Extrapolations to the basis set limits and electron correlation limits are performed. After this, we combine everything and report our best estimates of the final barriers of rotation. Finally, we discuss the equilibrium torsion angle.

3.1. Nonrelativistic Electronic Energies. We have calculated the torsional barriers over the planar, 0°, and the perpendicular, 90°, conformations, with various wave function methods. Several different basis sets have been used for single-point energy evaluations on the structures optimized at the B3LYP/TZVPP level. Full relaxation of the coordinates was allowed, with the exception of the torsional angle. In addition to the transition state structures, also the angle of the equilibrium structure was constrained, to the experimental value of 44.4°. A full optimization at the B3LYP level gives an angle of 39.5° at 0 K. The potential energy surface near the minimum is, however, very shallow, and the 44.4° conformation lies only 0.26 kJ/mol higher in energy. This is further discussed in Section 3.12, where thermal corrections are seen to have a large effect on the minimum angle.

Table 1. Computed Barriers of the Torsion around 0° and 90°, Using Selected Dunning Basis Sets[a]

	HF		MP2		SCS-MP2		CCSD		CCSD(T)	
	0°	90°	0°	90°	0°	90°	0°	90°	0°	90°
cc-pVDZ	12.82	5.40	12.23	7.68	12.24	6.73	11.34	6.56	10.89	7.23
aug-cc-pVDZ	12.05	4.57	9.86	7.45	10.15	6.45	9.77	6.08	9.23	6.67
cc-pVTZ	12.49	5.82	9.86	9.13	10.27	7.97	9.69	7.68	8.85	8.50
aug-cc-pVTZ	12.46	5.95	9.78	9.43	10.30	8.26	9.77	8.05	8.83	8.86
cc-pVQZ	12.56	5.88	9.65	9.33	10.14	8.13	9.64	7.90	8.68	8.74
aug-cc-pVQZ	12.53	5.81	9.35	9.31	9.89	8.12	9.41	7.92	8.39	8.76

[a] The barriers are given in kJ/mol.

Table 2. Barriers Computed at the HF and B3LYP Levels Using Fully Decontracted Polarization Consistent Basis Sets, As Well As Selected Karlsruhe Basis Sets[a]

	HF		B3LYP	
	0°	90°	0°	90°
pc-1	13.57	4.22	9.16	7.07
aug-pc-1	15.16	5.35	9.38	7.88
pc-2	12.70	5.62	8.04	8.32
aug-pc-2	12.52	5.71	7.75	8.40
pc-3	12.50	5.79	7.76	8.48
aug-pc-3	12.49	5.80	7.76	8.49
pc-4	12.47	5.81	7.76	8.50
SVP	11.71	6.04	5.92	9.57
TZVPP	12.55	5.75	7.73	8.40
QZVPP	12.53	5.79	7.83	8.47

[a] Energies are in kJ/mol.

Table 1 shows the barriers obtained with basis sets of increasing size. One can note that for all of the correlated wave function methods considered, the barrier at 90° seems to converge rather smoothly toward the aug-cc-pVQZ value, both when increasing the cardinal number of the basis set and when augmenting the basis with diffuse functions.

For the barrier around 0°, the situation is different. The barrier gets constantly lower compared to the perpendicular barrier, and, for CCSD(T), ΔE (0°) eventually falls below ΔE (90°), in agreement with the experimental results. However, there is still a notable lowering of the barrier when adding diffuse functions to the quadruple-ζ quality basis set, that is, when going from cc-pVQZ to aug-cc-pVQZ. At the CCSD(T) level, the difference is 0.3 kJ/mol. Therefore, the basis set limit cannot safely be considered to be reached.

We also want to stress, as has been done several times before, that correlated calculations at the cc-pVDZ level are highly unreliable, almost to the point of being useless. Little of the correlation energy inherent to a particular method is captured, and, what is worse, the amount is very different depending on conformation. As an example, the relative MP2 correlation energies at the cc-pVDZ level are −0.59 and +2.28 kJ/mol for the 0° and 90° barriers, respectively. This represents only 17% of the best estimate for the 0° conformation but 65% for the 90° conformation. Reasons for this, and the slower basis set convergence for the planar conformer in general, are discussed in more detail in connection with intramolecular basis set superposition error in Section 3.8.

3.2. Extrapolation toward the Basis Set Limit: Reference Energies. Using basis set extrapolation techniques, it is possible to obtain more accurate energies without performing prohibitively expensive calculations with larger basis sets. We begin by considering the reference Hartree−Fock energy.

A converged HF energy is naturally important, especially when very high accuracy is desired. With uncertainties in the reference energy, the incorporation of other, minute correction terms loses meaning. Although the correlation consistent series seems to be reasonably well converged also for the HF energy in Table 1, it is well-known that HF and DFT energies are not optimally represented by this series.[56,57] For this, we have employed Jensen's polarization-consistent basis sets, pc-*n*[26,27] and the augmented versions aug-pc-*n*.[28] The *n* in the basis set name indicates the polarization beyond the free atom. Thus, pc-1 for carbon is a double-ζ basis set with *s*, *p*, and *d* functions.

For the self-consistent field (SCF) energy extrapolations, both at the HF and B3LYP levels, we used two of the three-point schemes suggested by Jensen,[29] in connection with fully decontracted pc basis sets. Below, $L_{\max}$ is the highest angular momentum of the (carbon) basis set, N_s is the number of *s*-functions, and E_∞^{SCF}, B, and C are variables that need to be fitted using the energies of three consecutive pc-*n* basis sets:

$$E_\infty^{\mathrm{SCF}} \approx E_{L_{\max},N_s}^{\mathrm{SCF}} - B(L_{\max}+1)e^{-C\sqrt{N_s}} \tag{1}$$

The following, simpler formula, which does not take N_s as a parameter was also used:

$$E_\infty^{\mathrm{SCF}} \approx E_{L_{\max}}^{\mathrm{SCF}} - B(L_{\max})^{-C} \tag{2}$$

Table 2 shows the HF barriers computed with different pc basis sets; extrapolated values are found in Table 3. For comparison, we have also tested a few selected Karlsruhe basis sets[23,58,59] and in addition report B3LYP results.

The nice, smooth convergence of the barriers when climbing the pc-*n* ladder is noteworthy. The SCF energies, both for HF and B3LYP, are well converged at the pc-4 level. Both extrapolation formulas for pc-[2,3,4] give essentially the same barriers as the nonextrapolated value. The total energies are between 0.12 and 0.17 kJ/mol lower, though. The fitted C parameter for eq 1 is near 6 for all extrapolation combinations. This was noted in ref 29 and exploited for the construction of a two-point fitting scheme, the validity of which our results corroborate. For eq 2, the C parameter changes significantly when increasing $L_{\max}$. For both equations, the B parameter assumes very varied values.

Also for the SCF energies, the double-ζ basis sets pc-1, aug-pc-1, and SVP[58] give quite poor barriers. The Dunning (aug)-cc-pVDZ basis sets actually perform significantly better, see Table 1. The sometimes unsatisfactory perfor-

Table 3. Extrapolated Values for the Barriers Computed at the HF and B3LYP Levels Using Fully Decontracted Polarization Consistent Basis Sets[a]

	HF				B3LYP			
	0°	90°	$\tilde{B}$	$\tilde{C}$	0°	90°	$\tilde{B}$	$\tilde{C}$
				Using Eq 1				
pc-[1,2,3]	12.48	5.80	1.45×10^5	5.42	7.73	8.49	1.61×10^5	5.45
aug-pc-[1,2,3]	12.50	5.74	9.19×10^5	5.74	7.77	8.49	1.07×10^6	5.79
pc-[2,3,4]	12.46	5.82	1.02×10^6	6.04	7.77	8.50	1.32×10^6	6.12
				Using Eq 2				
pc-[1,2,3]	12.43	5.83	1.52×10^0	5.85	7.66	8.52	1.60×10^0	5.90
aug-pc-[1,2,3]	12.52	5.83	1.45×10^0	5.86	7.79	8.52	1.54×10^0	5.92
pc-[2,3,4]	12.46	5.82	5.17×10^3	11.31	7.77	8.50	6.20×10^3	11.45

[a] The average values of the fitting parameters *B* and *C* are also reported for each extrapolation series. Energies are in kJ/mol.

mance of pc-1 has been noted earlier.[56] It is also interesting to note that contrary to the correlated WF energies, augmenting the double-ζ basis set degrades the relative energies between the 44.4° and 0° conformations quite significantly. Possible reasons for this are discussed in Section 3.8.

From the data in Table 3, one cannot clearly recommend one extrapolation scheme over the other. When extrapolating the pc-*n*, $n = 2,3,4$ series, the formulations give very similar energies. Some differences using the smaller basis sets can be noted, but neither method consistently outperforms the other. Using eq 2, the total pc-[2,3,4] energies are *ca.* 0.05 kJ/mol lower. For this specific series, eq 2 was found to perform slightly better also for the smaller systems in ref 29. Probably due to its simpler formulation, the numerical solutions were also more stable, whereas solving the system of equations arising from eq 1 on occasion required manual fine-tuning of the initial guesses to converge. Also, although unambiguous for biphenyl, deciding on the values of N_s in eq 1 is not always trivial.[29] For these reasons, we use the values obtained by eq 2 as our reference energies.

3.3. Extrapolation toward the Basis Set Limit: Correlation Energies. With a reliable reference energy established, we now turn our attention to the correlation energy. We have used the two-point extrapolation scheme of Halkier et al.,[39] which has proven to be robust and reliable. Below, *X* and *Y* are the (consecutive) cardinal numbers of the two basis sets used in the extrapolation, that is, 3 for a cc-pVTZ basis, etc.:

$$E_{\infty}^{\text{corr}} \approx E_{XY}^{\text{corr}} = \frac{E_X^{\text{corr}} X^3 - E_Y^{\text{corr}} Y^3}{X^3 - Y^3} \qquad (3)$$

In Table 4, extrapolation combinations are given for the various choices of basis sets. The nonextrapolated aug-cc-pVQZ barriers from Table 1 are reproduced for convenience. The extrapolation corroborates the findings in Section 3.1: The relative energies between the 44.4° and 90° conformations are converged. Extrapolation using the aug-cc-pVTZ and aug-cc-pVQZ bases, aug-cc-pV[T,Q]Z, gives almost the same barriers as those of the nonextrapolated aug-cc-pVQZ basis set. But as suspected, there is still a significant lowering of the barrier at 0°, with augmentation by diffuse function being critical. Part of the difference between the pure aug-cc-pVQZ and the extrapolated values comes from the nonconverged HF reference energy. The last row in Table 4 shows the barriers calculated using the extrapolated reference HF energy together with the correlation energy of the aug-cc-pVQZ basis. The difference compared to the extrapolated correlated energies decreases but is still significant. The magnitude of the lowering is indeed so large, ~0.4 kJ/mol, that one cannot rule out a further lowering by employing even larger basis sets.

Table 4. Barriers of the Torsion around 0° and 90°, Using Extrapolated Values from Two Basis Sets[a]

	MP2		SCS-MP2		CCSD		CCSD(T)	
	0°	90°	0°	90°	0°	90°	0°	90°
cc-pV[D,T]Z	8.98	9.56	9.55	8.30	9.11	7.97	8.10	8.84
aug-cc-pV[D,T]Z	9.57	9.56	10.18	8.32	9.60	8.17	8.48	9.08
cc-pV[T,Q]Z	9.35	9.37	9.90	8.14	9.47	7.95	8.41	8.81
aug-cc-pV[T,Q]Z	8.93	9.34	9.47	8.13	9.03	7.94	7.96	8.79
aug-cc-pVQZ	9.35	9.31	9.89	8.12	9.41	7.92	8.39	8.76
aug-cc-pVQZ(erE)	9.29	9.32	9.82	8.13	9.34	7.93	8.33	8.77

[a] For example, the extrapolated value from the cc-pVTZ and cc-pVQZ basis sets is denoted cc-pV[T,Q]Z. The extrapolated pc-[2,3,4] values have been used as reference HF energies. For comparison, the raw, nonextrapolated aug-cc-pVQZ values are reported; (erE) denotes that the extrapolated reference energy has been used. Energies are in kJ/mol.

Slightly surprisingly, the extrapolated cc-pV[D,T]Z relative energies are, for all methods, quite close to the aug-cc-pV[T,Q]Z energies. Fortuitously, the energy differences between cc-pVDZ and cc-pVTZ results apparently reproduce the right convergence behavior. All extrapolated results get the barrier order for CCSD(T) correct, with ΔE (0°)<ΔE (90°). This underlines the usefulness of extrapolation; significantly better relative energies can be obtained compared to the nonextrapolated raw energies.

3.4. Comparison of Correlated Methods. Comparing the different hierarchies of correlated wave function methods to the reference CCSD(T) data, a few points can be observed:

a. The MP2 method overestimates ΔE (0°), ΔE (90°) by 1.0 and 0.6 kJ/mol, respectively, but gives the correct ordering of the barriers.

b. The SCS-MP2 method overestimates ΔE (0°) by 1.5 kJ/mol and underestimates ΔE (90°) by 0.7 kJ/mol, which leads to an erroneous ordering of the barriers.

c. The CCSD method overestimates ΔE (0°) by 1.1 kJ/mol and underestimates ΔE (90°) by 0.8 kJ/mol, which again leads to an erroneous ordering of the barriers.

Triple excitations on top of the CCSD energies are important. Table 5 shows the basis set dependence of the triples contribution, (T). The 90° values again seem well converged, while a slight increase in the absolute magnitude

Table 5. Perturbative Triples Contributions with Different Basis Sets and Extrapolation Combinations[a]

	0°	90°
cc-pVDZ	−0.45	+0.67
aug-cc-pVDZ	−0.54	+0.59
cc-pVTZ	−0.85	+0.82
aug-cc-pVTZ	−0.94	+0.81
cc-pVQZ	−0.96	+0.84
aug-cc-pVQZ	−1.02	+0.84
cc-pV[D,T]Z	−1.00	+0.88
aug-cc-pV[D,T]Z	−1.11	+0.91
cc-pV[T,Q]Z	−1.05	+0.86
aug-cc-pV[T,Q]Z	−1.07	+0.85

[a] Energies are in kJ/mol.

of the correction for the barrier at 0° is still present, when comparing the aug-cc-pVQZ and the aug-cc-pV[T,Q]Z results. The difference is however only 0.05 kJ/mol, much lower than the corresponding difference for the total barrier. This is in line with previous findings where the basis set dependence of the perturbative triples has been shown to be less severe than for CCSD single and double excitations.[60,61]

The triples correction contributes with different sign to the two barriers; the 0° barrier is lowered by ~1.1 kJ/mol, while the 90° barrier is raised by ~0.9 kJ/mol. This implies that triple excitations are the more abundant, the more planar the conformation is. A possible explanation for this is that as the planarity of the molecule increases, the electron density of *ortho*-hydrogens overlaps increasingly, giving more opportunities for triple excitations to occur.

The relatively good performance of MP2 is apparently rooted in the quite large triple excitation corrections when going from CCSD to CCSD(T). This is naturally something that cannot be represented at the second-order perturbation level, where only double excitations contribute to the energy. Thus the good agreement comes from a cancelation of errors. This also explains the failure of SCS-MP2, where the only difference to standard MP2 is that the same-spin and opposite-spin contributions to the correlation energy are scaled differently. In essence, no information about the triples contribution enters. SCS-MP2 was devised partly to damp the usual overestimation of long-range same-spin correlation in MP2.[33] In the case of biphenyl, this overestimation of MP2 fortuitously mimics the triples contribution. Thus, without artificially overestimated dispersion, the SCS-MP2 energies become worse and closer to the CCSD results.

The magnitude of the triples correction entices caution toward the adequacy of treating the triples in a perturbative manner, a full triples consideration might provide additional contributions to the relative energies. In addition to not going beyond perturbative triple excitations, CCSD(T) does not account for a possible multireference character present in the molecule (except indirectly, via the reasonably high percentage of correlation energy recovered). This is expected to be a minor omission, however. Diagnostics devised to quantify the reliability of a single-reference treatment support this view. The $T1$ diagnostic by Lee and Taylor[62] of the CCSD solution was in all cases found to be below 0.011. The $D1$ diagnostic by Janssen and Nielsen[63] was always below 0.030. Thus it appears quite safe to omit an explicit treatment of multiple reference configurations. However, as will be shown in Section 3.6, extrapolation toward the full configuration-interaction limit still has an appreciable effect on the barrier heights.

3.5. Core−Core and Core−Valence Correlation. To explore the error introduced by keeping the 1*s* orbital of carbon uncorrelated in the WF calculations, we performed calculations with all electrons correlated. This was done using the weighted core-valence basis sets of Peterson and Dunning.[38] Computational resources limited this study to the double- and triple-ζ basis sets, cc-pwCVDZ and cc-pwCVTZ, and subsequent extrapolation, although some cc-pwCVQZ calculations were performed, as discussed at the end of this section. The effect of correlating the core electrons in a given basis set was obtained by comparing energies with and without correlating the 1*s* electrons of the carbons. Table 6 shows that the relative corrections that core-correlation introduce are small but nonvanishing. The extrapolation was performed with eq 3 using only the core-correlation contribution, *not* the full correlation energy. The extrapolated cc-pwCV[D,T]Z values are arguably, with the assumption that the core correlation energy follows the same convergence pattern as the total correlation energy, the most accurate. For all methods, core correlation is thus seen to raise both barriers slightly, the effect on the 0° conformation being more pronounced.

The double-ζ cc-pwCVDZ basis set is again much too small to give reasonable results for any of the correlated WF methods. Even if the correction is small, its magnitude and even its sign change when using larger basis sets. Thus, for estimating core-correlation effects the use of at least a triple-ζ basis is mandatory, lest the "correction" turns to degradation.

The basis set problems are again more severe for the planar conformation, while the convergence of the 90° barrier is much smoother. Thus the uncertainties in the core-correlation corrections for the 0° barrier are bigger than for the perpendicular; the 0° corrections should probably be even slightly more positive, i.e., raise the barrier slightly more. To test this, we performed calculations with the quadruple-ζ cc-pwCVQZ basis set. This basis set was too large for fully correlated computations on the 44.4° conformer with available resources, so only the relative corrections between the planar and perpendicular conformers could be obtained. Comparing the barriers obtained at the cc-pwCV[D,T]Z and cc-pwCV[T,Q]Z levels, it was found that the 0° barrier is indeed slightly raised when using the more complete extrapolation: For MP2 and SCS-MP2 by 0.02 kJ/mol and for CCSD and CCSD(T) by 0.03 kJ/mol, compared to the 90° conformer. As the core-correlation contribution to the 90° barrier seems quite converged already at the cc-pwCV[D,T]Z level, much of this would likely be transferable also to the relative energies between the 0° and 44.4° conformations.

Also for core-correlation, the two-point extrapolation scheme, eq 3, at least in this case, captures more of the correlation energy than the raw-values of a basis set one step ahead in the series. The relative energies are of comparable accuracy, that is, cc-pwCV[D,T]Z gives essentially the same relative energies as cc-pwCVQZ. The relative core-correlation at the CCSD(T) level is also almost the same as for

Table 6. Relative Core-Correlation Corrections to the Torsion around 0° and 90°, Using the Weighted Core-Correlation Basis Sets[a]

	MP2		SCS-MP2		CCSD		CCSD(T)	
	0°	90°	0°	90°	0°	90°	0°	90°
cc-pwCVDZ	−0.00	0.02	−0.06	0.01	−0.03	0.03	−0.04	0.02
cc-pwCVTZ	0.09	0.03	0.08	0.02	0.05	0.04	0.04	0.03
cc-pwCV[D,T]Z	0.13	0.04	0.11	0.02	0.09	0.05	0.08	0.03

[a] cc-pwCV[D,T]Z again denotes extrapolated values. Energies are in kJ/mol.

Table 7. Barriers of the Torsion around 0° and 90°, Calculated with the Continued-Fraction CC Method (CC-cf), Using the Best Estimates of the Total Energies at the HF, CCSD, and CCSD(T) Levels[a]

	0°	90°
CC-cf	7.88	8.94
CCSD(T)	8.04	8.83

[a] Also shown are CCSD(T) barriers including core-correlation. Energies are in kJ/mol.

CCSD, indicating that triple excitations from the core are not significant for the relative energies. A more complete dissertation of the cc-pwCVQZ results can be found in the Supporting Information.

3.6. Extrapolation toward the Full Configuration-Interaction Limit. With the best estimates for the electronic energies at different levels of theory available, we next proceed to extrapolation of the coupled cluster series toward completeness. For this, we have employed the continued fraction method of Goodson (CC-cf),[40] devised to provide near full configuration-interaction (FCI) energies. Below, $\delta_1 = E$ [HF], $\delta_2 = E$[CCSD] $-$ E[HF], and $\delta_3 = E$[CCSD(T)] $-$ E[CCSD]:

$$E[\text{CC-cf}] = \frac{\delta_1}{1 - \dfrac{\delta_2/\delta_1}{1 - \delta_3/\delta_2}} \tag{4}$$

For HF, the pc-[2,3,4] energies were used; for CCSD and CCSD(T), the aug-cc-pV[T,Q]Z energies with cc-pwCV[D,T]Z core-correlation contributions added were used. We note that all conformations were treated individually, after which the relative energies were computed.

Table 7 shows the barriers as obtained with eq 4 and represents our best estimates for the nonrelativistic electronic energies at 0 K. Compared to the CCSD(T) results, the barrier at 0° is lowered slightly more, while the 90° barrier is raised.

3.7. The Effect of Relativity. To check, and possibly rule out the effect of relativity on the barrier heights, we have performed relativistic all-electron calculations at the B3LYP level, using the fully decontracted TZVPP basis set, treated as Cartesian. Table 8 shows the barriers at the nonrelativistic Lévy-Leblond level[64] and using a one-step exact two-component relativistic Hamiltonian (X2C).[42]

The barriers are seen to be virtually unaffected by relativity. Even with the speed of light artificially halved to 0.5c, only a minute enhancement of the relativistic corrections is observed. An even more rigorous treatment of relativity, beyond X2C, might still have a small effect on the relative energies, but, for practical purposes, relativity can safely be considered not to contribute to the barrier heights in biphenyl.

Table 8. Barriers of the Torsion around 0° and 90°, Calculated at the B3LYP Density Functional Level, Using the Decontracted TZVPP Basis Set, at Nonrelativistic Lévy-Leblond (NR) and One-Step Exact Two-Component Relativistic (X2c) Levels[a]

	0°	90°
NR	7.98	8.39
X2C	7.99	8.39
X2C(0.5c)	7.83	8.38

[a] (0.5c) denotes calculations done with the speed of light halved. Energies are in kJ/mol.

3.8. Intramolecular Basis Set Superposition Error. As discussed above, the convergence toward the basis set limit is much slower for the planar conformation compared to that of the perpendicular. This has previously been attributed to the more demanding basis set requirement for describing the dispersion interaction in planar biphenyl.[11] In this Section, our working hypothesis will be that much of the difference instead arises from intramolecular basis set superposition error (BSSE).

Intramolecular BSSE is more difficult to assess than intermolecular BSSE between two fragments,[65−67] where the counterpoise (cp) correction scheme[68] has become a *de facto* standard. Jensen used an approach analogous to cp in a study of the BSSE for relative energies between different conformers of the same molecule.[66] When comparing the relative energies between the conformers, the logical suggestion was to explore the BSSE by the combined basis set of both conformers, that is, by inserting dummy, ghost atomic centers at the positions the other conformer would occupy, where the conformations superimposed. In the spirit of Jensen's method, we will compare these values with the results obtained in the normal basis sets, consisting of functions only on the atoms.

For simplicity, we have only considered the relative energies between the planar and perpendicular conformations. With this combination, the basis sets are augmented by 8 ghost centers, corresponding to the four carbons and four hydrogens that would stick out of the plane of one of the phenyl rings, if the structures would be superimposed. Figure 2 shows the situation for the planar conformer. Although the optimized geometries were used in the calculations, we have assumed that the rest of the atoms are located at the same relative positions in the two conformers, so as not to add dummy centers that nearly coincide with the atoms. The calculations were performed in C_{2v} symmetry.

Table 9 shows the relative energies, with and without corrections for BSSE. For the double-ζ basis sets, the addition of the ghost centers leads to a significant lowering of the 0° energy compared to the 90°. When going to the

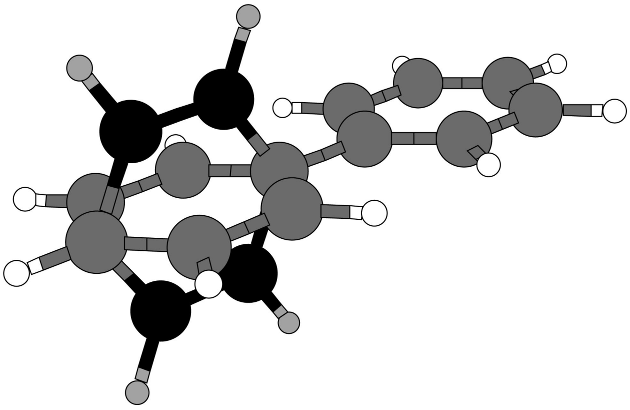


Figure 2. The planar conformation of biphenyl, with the ghost atom centers defined by the perpendicular conformation shown in black and gray.

Table 9. Relative Energies between the 0° and 90° Conformations, Calculated with Selected Basis Sets[a]

	HF	MP2	SCS-MP2	CCSD	CCSD(T)
cc-pVDZ	7.42	4.55	5.51	4.78	3.66
cc-pVDZ(cp)	6.61	1.67	3.20	3.00	1.42
Δ(cp)	−0.81	−2.88	−2.31	−1.78	−2.24
aug-cc-pVDZ	7.48	2.40	3.70	3.63	3.32
aug-cc-pVDZ(cp)	6.05	−0.27	1.38	1.21	−0.45
Δ(cp)	−1.43	−2.67	−2.32	−2.42	−3.77
cc-pVTZ	6.66	0.73	2.30	2.00	0.35
cc-pVTZ(cp)	6.62	0.58	2.25	1.92	0.15
Δ(cp)	−0.04	−0.15	−0.05	−0.08	−0.20
cc-pV[D,T]Z	-	−0.58	1.25	1.14	−0.74
cc-pV[D,T]Z(cp)	-	0.13	1.87	1.49	−0.37
Δ(cp)	-	+0.71	+0.62	+0.35	+0.37
aug-cc-pVQZ	6.72	0.04	1.77	1.49	−0.37
aug-cc-pV[T,Q]Z	-	−0.41	1.34	1.09	−0.83

[a] (cp) denotes that the basis set has been augmented by functions at the ghost centers of the other conformation. Δ(cp) is the difference between the normal and the (cp) basis sets. Energies are in kJ/mol; a positive value denotes that the 0° energy is higher.

cc-pVTZ basis set, things look better, and the difference between corrected and uncorrected energies becomes reasonably small.

An anomaly can be found in the extrapolated cc-pV[D,T]Z values, where the cp corrected energy differences are much further from the best estimates compared to the noncorrected ones, even though the cp-corrected basis sets are more complete. This also shows up as an artificially large BSSE correction, Δ(cp), which even has the "wrong" sign. The underlying reason is the fortuitously good performance of the cc-pV[D,T]Z energies, as discussed in Section 3.3; the extrapolated cc-pV[D,T]Z(cp) values are better than the raw cc-pVTZ(cp) values, as they are expected to be.

It is difficult to divide the energy differences between the normal and cp-augmented basis sets into components arising from just a larger flexibility of the basis set and BSSE. The fact that the difference decreases significantly when going from double-ζ to larger basis sets does suggest that a major part in fact is due to BSSE. A possible explanation for the origin of BSSE, which favors the 90° conformation over the 0° conformation, could be the following. In the planar conformation, basis functions are present only in the plane of the molecule. For the perpendicular conformation, the centers are naturally present in all three dimensions. Thus, in the perpendicular case, the phenyl planes can to an extent utilize basis functions from the other plane to describe the space above (and below) their own plane. This suggestion is supported by the fact that, for all methods save MP2, the correction term is bigger for aug-cc-pVDZ than cc-pVDZ: With augmented diffuse functions on the centers, they extend more efficiently over to the top (and bottom) of the other plane. This would explain the much slower convergence of the energy in the planar conformer, where the space above the planes has to be described only by basis functions in the plane. For the same reason, the performance of the normal augmented double-ζ basis sets compared to the nonaugmented is poorer, as seen also for the pc-1 basis set in Section 3.2.

Much of the poor performance of the smaller basis sets in describing the relative energy of the planar and twisted conformers thus comes not from an intrinsically poorer description of the planar system but from a better possibility of the twisted conformers to "borrow" basis functions from the opposite plane. From Table 9, one can also note the usual observation that the BSSE is more pronounced for the correlated WF methods compared to HF.

3.9. Zero-Point Vibrational Energy and Thermal Corrections. In this section we consider the zero-point vibrational energy (ZPE) contributions to the barriers. The experimental values were measured at approximately the nozzle temperature of 401 K,[6] while the reported values for the potential barriers refer to absolute zero.[5] Nevertheless, it is of interest to explore the temperature dependence of the energetics, so we have also explored finite temperature effects on the barriers.

We consider the vibrational contributions to the thermal energy corrections with two different approaches. The basis for this is the observation that internal coordinate analysis shows the lowest vibrational frequency consistently to correspond almost purely to the internal rotation of the phenyl planes. This frequency becomes imaginary sufficiently far from equilibrium and naturally is imaginary also at the transition states at 0° and 90° angles. When calculating the ZPE, imaginary frequencies do not contribute to the sum over vibrations, lowering the ZPE. In general, the imaginary vibrations of transition states do not directly match a vibration in the ground-state geometry, but, in the special case here, there is a one-to-one correspondence. Therefore, it might be motivated to remove the lowest vibration also from the ZPE and vibrational contribution to the enthalpy (H) and free energy (G). A similar approach was discussed previously by Dos Santos et al.[69] We leave this option to the reader, but, subsequently in this work, we will consider the traditional method of including all real vibrations. Another approach would be to treat the internal rotation of the phenyl planes as a hindered rotation, as the barrier heights of *ca.* 8 kJ/mol are comparable to k_BT, 3.3 kJ/mol at experimental temperature.

In Table 10, the ZPE as well as relative enthalpies and free energies at the experimental temperature are shown, using both approaches discussed above, obtained at the B3LYP/TZVPP level. The ZPE has a notable effect on

Table 10. Lowest Frequencies (ν_1), Zero-Point Energies (ZPE), Relative Enthalpies (ΔH), and Free Energies (ΔG) for Biphenyl with Different Torsional Angles[a]

	all real freqs			3*N*−7 highest freqs		
	0°	44.4°	90°	0°	44.4°	90°
ν_1	80.5 *i*	61.0	55.3 *i*			
ZPE	475.80	475.68	475.02	475.80	475.32	475.02
ΔZPE	+0.12	0	−0.66	+0.49	0	−0.30
Δ*H* (401 K)	518.14	521.04	517.58	518.14	517.69	517.58
ΔΔ*H*	−2.90	0	−3.46	+0.45	0	−0.12
Δ*G* (401 K)	352.28	345.92	349.98	352.28	350.98	349.98
ΔΔ*G*	+6.36	0	+4.06	+1.30	0	−1.00

[a] Values calculated using all real frequencies are shown together with values calculated using all except the lowest frequency. Calculated at the B3LYP/TZVPP level. Frequencies are in cm^{-1}; energies are in kJ/mol.

the relative energies. As expected, the barriers are raised if the lowest vibration is consistently omitted. This is especially pronounced in the case of enthalpy and free energy corrections. Without the frequency of the internal rotation, there is a very small change between the relative zero point energies and the enthalpies, indirectly confirming that other vibrational modes are largely independent of the rotation.

It should be noted that we have chosen not to scale the vibrational frequencies, a common procedure used to approximate the effects of anharmonicity. In this case, it is not clear that scaling would provide an overall improvement of the frequencies, so adding an extra empirical scaling factor is not really motivated. Some inherent uncertainty in the computed zero-point energies and the thermal corrections based on the vibrations thus exists. Compared to the other remaining sources of uncertainty, this is probably the most significant.

3.10. The Effect of Geometry on the Relative Energies. All energetics discussed in previous sections are based on single-point evaluations on geometries optimized at the density functional B3LYP level. Although expected to be highly realistic, it is difficult to estimate exactly how good the geometries are, without performing full optimizations at a sufficiently high *ab initio* level. This is however, for the time being, beyond reasonable computational resources. For this reason, it is of interest to examine how sensitive the relative energies are with respect to the exact geometries of the 0°, 44.4°, and 90° conformations.

To explore this, we have reoptimized the geometries at Hartree−Fock and the RI-MP2 level, with subsequent single point energy calculations. The geometry optimizations used the same basis set as for the B3LYP geometries, i.e., TZVPP. Table 11 shows the relative cc-pVTZ energies at different levels of theory, based on geometries optimized at the B3LYP, MP2, and HF levels.

Some differences in the barriers can be noted, although none are very large. Even the barriers based on HF geometries deviate less than 0.2 kJ/mol in the case of correlated WF methods. The differences are largest for the HF and MP2 energies, where for the specific method more optimal geometries apparently play an especially prominent role. Reassuringly, the coupled cluster barriers are very similar for the electron correlation including B3LYP and MP2 geometries. We note that the total CCSD(T)/cc-pVTZ energies are *ca.* 0.5 kJ/mol lower when using the MP2 geometries instead of the B3LYP geometries, suggesting the MP2 geometries to be slightly superior. Again, the planar conformation exhibits the largest sensitivity, also with respect to geometries. However, all-in-all, even more accurate geometries are expected to have a very minor effect on the relative barriers.

3.11. Best Estimates of the Torsional Barriers. With values for all contributions to the barriers at zero temperature available, we are now able to sum them up. Table 12 summarizes the work. The best estimate values of the barriers are $\Delta E(0°) = 8.0$ kJ/mol, and $\Delta E(90°) = 8.3$ kJ/mol. Thus, we have succeeded in bringing the barriers to within the reported experimental uncertainty, reconciling theory with experiment. Although some uncertainties remain in the calculated values, the most important being *(i)* the zero-point energy and *(ii)* the remaining basis-set effect for $\Delta E(0°)$, we are fairly confident that the experimental values of 6.0 ± 2.1 and 6.5 ± 2.0 kJ/mol for the potential energy at 0° and 90°, respectively, are at the low end.

3.12. The Experimental vs Computed Torsional Angle. Finally, we turn to the equilibrium torsional angle. To pinpoint the exact minimum, the potential energy was calculated at 1° intervals between 33° and 51°. The correlated WF energies were obtained with the aug-cc-pVDZ and aug-cc-pVTZ basis sets, with subsequent extrapolation (aug-cc-pV[D,T]Z) as discussed in previous sections. The Hartree−Fock reference energies and B3LYP energies were obtained with the doubly polarized quadruple-ζ QZVPP[59] basis set. On top of the 0 K energies, thermal corrections were added, based on frequency calculations at the B3LYP/TZVPP level. The obtained energies were then fitted to a harmonic function of the usual form $E(\phi) = a \times (\phi\text{-}\phi_0)^2 + c$, with a, ϕ_0, and c free parameters, and ϕ_0 the resulting minimum torsion angle. To ensure that the region of the fit was as harmonic as possible, the final fit was performed using only points within 6° of the minimum angle, that is, using the 12 closest data points. Limiting the number of data points changed the angle very little, on average less than 0.1°, but resulted in slightly better fits.

Table 13 shows the calculated equilibrium angles at different levels of theory. The energy difference between the minimum angle and the 44.4° conformations, based on the fitted harmonic functions, are also shown. In Figure 3, the resulting curves for the electronic energy at the CC-cf, CCSD(T), and B3LYP levels are drawn. All correlated methods are in close agreement, giving a minimum angle of 38.8−41.0°, while the Hartree−Fock angle is much larger. The values and the curves are slightly shifted relative to each other, consistent with the relative energies of the 0° and 90° barriers at each level of theory. As for the energies, one can note that MP2 agrees quite well with the CC-cf and CCSD(T) results, both concerning the angle and the difference to the 44.4° conformation; SCS-MP2 again degrades both numbers. Also B3LYP fares well in the competition.

The experimentally measured[6] torsion angle of 44.4 ± 1.2° is significantly larger than the theoretical values at zero temperature; finite temperature effects can be expected to

Table 11. Torsional Barriers Based on Geometries Obtained at the B3LYP, MP2, and HF Levels of Theory, Using the cc-pVTZ Basis Set[a]

	HF		MP2		SCS-MP2		CCSD		CCSD(T)	
optimization level	0°	90°	0°	90°	0°	90°	0°	90°	0°	90°
B3LYP	12.49	5.82	9.86	9.13	10.27	7.97	9.69	7.68	8.85	8.50
MP2	12.27	5.78	9.99	9.11	10.32	7.94	9.65	7.69	8.86	8.52
HF	12.23	5.95	10.00	9.00	10.37	7.90	9.78	7.69	9.04	8.45

[a] Energies are in kJ/mol.

Table 12. Torsional Barriers at 0 K, Including All Corrections Considered: Nonrelativistic Electronic Energy (*E*(0 K), Including Core-Correlation), Relativistic Correction (RC), and Zero-Point Energy (ZPE)[a]

	CC-cf		CCSD(T)		B3LYP	
	0°	90°	0°	90°	0°	90°
E(0 K)	7.878	8.942	8.037	8.825	7.765	8.498
RC	+0.013	−0.002	+0.013	−0.002	+0.013	−0.002
ZPE	+0.120	−0.664	+0.120	−0.664	+0.120	−0.664
total	8.01	8.28	8.17	8.16	7.90	7.83

[a] The best-estimate continued-fraction barriers (CC-cf) are compared to those of CCSD(T) and B3LYP. Energies are in kJ/mol.

Table 13. Equilibrium Angle of Biphenyl Calculated at Different Levels: Electronic Energy at 0 K, and with Corrections for Zero-Point Energy (ZPE), Enthalpy (ΔH), and Free Energy (ΔG) at 401 K (CC-cf/B3LYP Level Only)[a]

	$\angle_{\min}$	ΔE(44.4°)
	Electronic Energy at 0 K	
HF	45.5°	0.01
MP2	39.6°	0.27
SCS-MP2	41.0°	0.13
CCSD	40.6°	0.15
CCSD(T)	39.0°	0.32
CC-cf	38.8°	0.36
B3LYP	39.4°	0.26
	CC-cf with Correction Terms	
ZPE	40.0°	0.20
ΔH	39.6°	0.27
ΔG	45.8°	0.01

[a] The energy differences to the 44.4° angle conformation are also given, in kJ/mol.

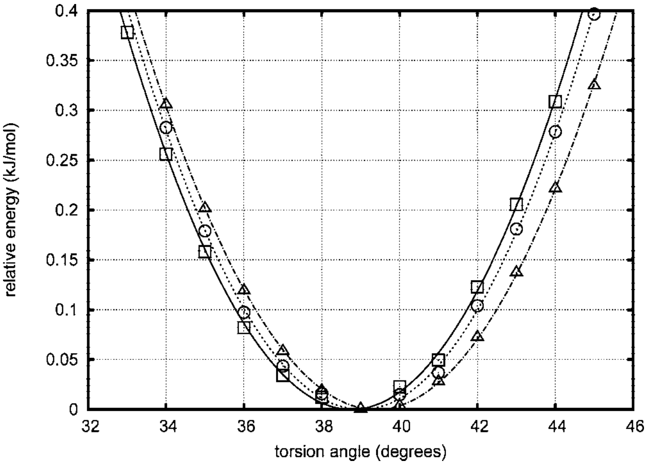


Figure 3. Potential energy curves near equilibrium, based on electronic energies at 0 K, calculated at the CC-cf (boxes), CCSD(T) (circles), and B3LYP (triangles) levels. The curves represent the harmonic potential fitted to the individual data points.

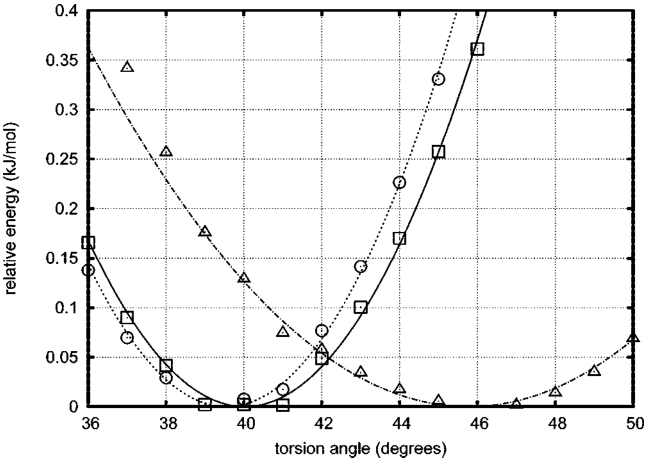


Figure 4. Potential energy curves near equilibrium, calculated at the CC-cf/B3LYP level. The electronic energies have been corrected for zero-point energy (boxes), enthalpy (circles), and free energy (triangles) at 401 K. The curves represent the harmonic potential fitted to the individual data points.

be important. Figure 4 shows the potential energy curve calculated at the CC-cf level, with thermal corrections at the B3LYP level added. From Table 13 one sees that zero-point energy and enthalpy corrections increase the angle but only by *ca.* 1°. Accounting for entropy effects via the free energy ΔG, on the other hand, increases the angle significantly, by 7.0° to 45.8°.

Even with free energy accounted for, the computed equilibrium angle lies outside the experimental error bars, being slightly larger than the measured angle. The approximations used in the computations should be kept in mind. As discussed in relation with the relative energies between the 0° and 90° conformations, using larger basis sets would shift the curves slightly toward a lower angle, most likely nudging the equilibrium angle to within experimental error. Although this would bring the calculated angle in agreement with experiment, a probably larger source of error is the free energy correction itself. This has been approximated via the vibrational frequencies that themselves contain uncertainties as discussed in Section 3.9. The effect of these approximations on the angle is difficult to predict, and a shift in either direction is possible. Further, all the energy curves are very shallow around the minimum, especially so the free energy surface. Small changes in the energies would have a notable effect on the angle. We are however fairly confident that the zero temperature angle is much smaller than the experimentally measured one. To reproduce the measured angle, entropy and free energy are thus seen to be compulsory ingredients of the computational recipe.

4. Conclusions and Outlook

We have reported accurate calculations of the barriers of rotation around the central bond in biphenyl. In contrast to previous studies, we have shown that the experimentally inferred barriers *can* be reproduced, if sufficiently sophisticated methodology is employed. Our calculated best estimates for the barriers are $\Delta E(0°) = 8.0$ and $\Delta E(90°) = 8.3$ kJ/mol, while the values estimated from experimental data[5] were $\Delta E(0°) = 6.0 \pm 2.1$ and $\Delta E(90°) = 6.5 \pm 2.0$ kJ/mol. Although the calculated barriers fit within the reported experimental uncertainty, with the same order, the expected accuracy of our results strongly suggests that the true values are close to the upper limit of the error bars reported.

The basis set dependence of the barriers, especially around 0°, is high. Even augmented quadruple-ζ basis sets are insufficient; the use of extrapolation techniques are compulsory. This is the case also for the electron correlation treatment, where an extrapolation toward the full configuration-interaction limit is necessary, as the coupled cluster CCSD(T) approach still has a bias toward twisted conformations.

The equilibrium angle of biphenyl was investigated, and free energy corrections were found to be necessary to widen the computed angle toward the experimentally measured value[6] of 44.4 ± 1.2°. Due to the shallowness of the potential near minimum, the calculated value of 45.8° is, however, very sensitive to small errors and uncertainties in the computed energies.

Our results can readily serve as a basis for future benchmark calculations for various theoretical treatments. The reported standard, canonical CCSD(T) results, obtained with basis sets up to the augmented quadruple-ζ level (aug-cc-pVQZ), are among the largest performed to date. The nature of biphenyl, having a highly delocalized electronic structure, should prove an interesting test case for, e.g., local *ab initio* correlation methods and other approximate methods aiming toward a lower computational scaling. To this end, we provide the structures and total energies as Supporting Information.

For the computationally less demanding density functional approach, the intricate competition between different chemical interactions makes for a delicious challenge, especially for nonempirical functionals that have to rely on a good description of the chemistry and physics involved, without the help of fitted parameters. Indirectly, the B3LYP functional was already benchmarked here and found to agree well with the highest-level extrapolated *ab initio* results and experiment. Its good performance could be considered somewhat fortuitous, however, due to the empirical nature of the functional.

Acknowledgment. We are grateful to Prof. Svein Samdal, Dr. Frank Jensen, and Dr. Michael Patzschke for discussions. CSC—The Finnish IT Center for Science—provided abundant computational resources.

Supporting Information Available: Details, atomic coordinates, and total energies. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Matta, C. F.; Hernandéz-Trujillo, J.; Tang, T.-H.; Bader, R. F. W. Hydrogen-Hydrogen Bonding: A Stabilizing Interaction in Molecules and Crystals. *Chem. Eur. J.* **2003**, *9*, 1940–1951.

(2) Poater, J.; Solà, M.; Bickelhaupt, F. M. Hydrogen-Hydrogen Bonding in Planar Biphenyl, Predicted by Atoms-In-Molecules Theory, Does Not Exist. *Chem. Eur. J.* **2006**, *12*, 2889–2895.

(3) Pacios, L. F. A theoretical study of the intramolecular interaction between proximal atoms in planar conformations of biphenyl and related systems. *Struct. Chem.* **2007**, *18*, 785–795.

(4) Hodges, M. P. XMakemol: a program for visualizing atomic and molecular systems, version 5, 2001. See http://www.non-gnu.org/xmakemol/(accessed June 19, 2008).

(5) Bastiansen, O.; Samdal, S. Structure and barrier of internal rotation of biphenyl derivatives in the gaseous state. Part 4. Barrier of internal rotation in biphenyl, perdeuterated biphenyl and seven non-ortho-substituted halogen derivatives. *J. Mol. Struct.* **1985**, *128*, 115–125.

(6) Almenningen, A.; Bastiansen, O.; Fernholt, L.; Cyvin, B. N.; Cyvin, S. J.; Samdal, S. Structure and barrier of internal rotation of biphenyl derivatives in the gaseous state Part 1. The molecular structure and normal coordinate analysis of normal biphenyl and perdeuterated biphenyl. *J. Mol. Struct.* **1985**, *128*, 59–76.

(7) Häfelinger, G.; Regelmann, C. Refined ab initio 6−31G split-valence basis set optimization of the molecular structures of biphenyl in twisted, planar, and perpendicular conformations. *J. Comput. Chem.* **1987**, *8*, 1057–1065.

(8) Tsuzuki, S.; Tanabe, K. Ab Initio Molecular Orbital Calculations of the Internal Rotational Potential of Biphenyl Using Polarized Basis Sets with Electron Correlation Correction. *J. Phys. Chem.* **1991**, *95*, 139–144.

(9) Rubio, M.; Merchán, M.; Ortí, E. The internal rotational barrier of biphenyl studied with multiconfigurational second-order perturbation theory (CASPT2). *Theor. Chim. Acta* **1995**, *91*, 17–29.

(10) Karpfen, A.; Choi, C. H.; Kertesz, M. Single-Bond Torsional Potentials in Conjugated Systems: A Comparison of ab Initio and Density Functional Results. *J. Phys. Chem. A* **1997**, *101*, 7426–7433.

(11) Tsuzuki, S.; Uchimaru, T.; Matsumura, K.; Mikami, M.; Tanabe, K. Torsional potential of biphenyl: Ab initio calculations with the Dunning correlation consisted basis sets. *J. Chem. Phys.* **1999**, *110*, 2858–2861.

(12) Arulmozhiraja, S.; Fujii, T. Torsional barrier, ionization potential, and electron affinity of biphenyl–A theoretical study. *J. Chem. Phys.* **2001**, *115*, 10589–10594.

(13) Grein, F. Twist Angles and Rotational Energy Barriers of Biphenyl and Substituted Biphenyls. *J. Phys. Chem. A* **2002**, *106*, 3823–3827.

(14) Grein, F. New theoretical studies on the dihedral angle and energy barriers of biphenyl. *J. Mol. Struct. (Theochem)* **2003**, *624*, 23–38.

(15) Grein, F. Influence of diffuse and polarization functions on the second-order Møller-Plesset optimized dihedral angle of biphenyl. *Theor. Chem. Acc.* **2003**, *109*, 274–277.

(16) Sancho-García, J. C.; Cornil, J. Anchoring the Torsional Potential of Biphenyl at the ab Initio Level: The Role of Basis Set versus Correlation Effects. *J. Chem. Theory Comput.* **2005**, *1*, 581–589.

(17) Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* **1964**, *136*, B864–B871.

(18) Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965**, *140*, A1133–A1138.

(19) Perdew, J. P.; Wang, Y. Accurate and simple density functional for the electronic exchange energy: Generalized gradient approximation. *Phys. Rev. B* **1986**, *33*, 8800–8802.

(20) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

(21) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785–789.

(22) 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.

(23) 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.

(24) Hartree, D. R. The wave mechanics of an atom with a non-coulomb central field. I. Theory and methods. *Proc. Cambridge Phil. Soc.* **1928**, *25*, 89–110.

(25) Fock, V. Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems. *Z. Phys.* **1930**, *61*, 126–148.

(26) Jensen, F. Polarization consistent basis sets: Principles. *J. Chem. Phys.* **2001**, *115*, 9113–9125.

(27) Jensen, F. Polarization consistent basis sets. II. Estimating the Kohn-Sham basis set limit. *J. Chem. Phys.* **2002**, *116*, 7372–7379.

(28) Jensen, F. Polarization consistent basis sets. III. The importance of diffuse functions. *J. Chem. Phys.* **2002**, *117*, 9234–9240.

(29) Jensen, F. Estimating the Hartree-Fock limit from finite basis set calculations. *Theor. Chem. Acc.* **2005**, *113*, 267–273.

(30) Møller, C.; Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev.* **1934**, *46*, 618–622.

(31) Weigend, F.; Häser, M. RI-MP2: first derivatives and global consistency. *Theor. Chem. Acc.* **1997**, *97*, 331–340.

(32) Weigend, F.; Häser, M.; Patzelt, H.; Ahlrichs, R. RI-MP2: optimized auxiliary basis sets and demonstration of efficiency. *Chem. Phys. Lett.* **1998**, *294*, 143–152.

(33) Grimme, S. Improved second-order Møller-Plesset perturbation theory by separate scaling of parallel- and antiparallel-spin pair correlation energies. *J. Chem. Phys.* **2003**, *118*, 9095–9102.

(34) Purvis, G. D., III; Bartlett, R. J. A full coupled-cluster singles and doubles model: The inclusion of disconnected triples. *J. Chem. Phys.* **1982**, *76*, 1910–1918.

(35) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. A fifth-order perturbation comparison of electron correlation theories. *Chem. Phys. Lett.* **1989**, *157*, 479–483.

(36) Dunning, T. H., Jr. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007–1023.

(37) Kendall, R. A.; Dunning, T. H., Jr; Harrison, R. J. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *J. Chem. Phys.* **1992**, *96*, 6796–6806.

(38) Peterson, K. A.; Dunning, T. H., Jr. Accurate correlation consistent basis sets for molecular core-valence correlation effects: The second row atoms Al-Ar, and the first row atoms B-Ne revisited. *J. Chem. Phys.* **2002**, *117*, 10548–10560.

(39) Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. Basis-set convergence in correlated calculations on Ne, N_2, and H_2O. *Chem. Phys. Lett.* **1998**, *286*, 243–252.

(40) Goodson, D. Z. Extrapolating the coupled-cluster sequence toward the full configuration-interaction limit. *J. Chem. Phys.* **2002**, *116*, 6948–6956.

(41) Deglmann, P.; Furche, F.; Ahlrichs, R. An efficient implementation of second analytical derivatives for density functional methods. *Chem. Phys. Lett.* **2002**, *362*, 511–518.

(42) Iliaš, M.; Saue, T. An infinite-order two-component relativistic Hamiltonian by a simple one-step transformation. *J. Chem. Phys.* **2007**, *126*, 064102.

(43) Werner, H.-J. et al. MOLPRO, version 2006 1, a package of ab initio programs, 2006. See http://www.molpro.net (accessed June 19, 2008).

(44) Hampel, C.; Peterson, K. A.; Werner, H.-J. A comparison of the efficiency and accuracy of the quadratic configuration interaction (QCISD), coupled cluster (CCSD), and Brueckner coupled cluster (BCCD) methods. *Chem. Phys. Lett.* **1992**, *190*, 1–12.

(45) Deegan, M. J. O.; Knowles, P. J. Perturbative corrections to account for triple excitations in closed and open shell coupled cluster theories. *Chem. Phys. Lett.* **1994**, *227*, 321–326.

(46) Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. Electronic structure calculations on workstation computers: The program system Turbomole. *Chem. Phys. Lett.* **1989**, *162*, 165–169.

(47) Häser, M.; Ahlrichs, R. Improvements on the direct SCF method. *J. Comput. Chem.* **1989**, *10*, 104–111.

(48) Treutler, O.; Ahlrichs, R. Efficient molecular numerical integration schemes. *J. Chem. Phys.* **1995**, *102*, 346–354.

(49) von Arnim, M.; Ahlrichs, R. Geometry optimization in generalized natural internal coordinates. *J. Chem. Phys.* **1999**, *111*, 9183–9190.

(50) Hättig, C.; Weigend, F. CC2 excitation energy calculations on large molecules using the resolution of the identity approximation. *J. Chem. Phys.* **2000**, *113*, 5154–5161.

(51) Hättig, C.; Hellweg, A.; Köhn, A. Distributed memory parallel implementation of energies and gradients for second-order Møller-Plesset perturbation theory with the resolution-of-the-identity approximation. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1159–1169.

(52) Jensen, H. J. A. et al. Dirac, a relativistic ab initio electronic structure program, Release DIRAC08.beta, 2008. See http://dirac.chem.sdu.dk (accessed June 19, 2008).

(53) Eichkorn, K.; Weigend, F.; Treutler, O.; Ahlrichs, R. Auxiliary basis sets for main row atoms and transition metals and their use to approximate Coulomb potentials. *Theor. Chem. Acc.* **1997**, *97*, 119–124.

(54) Feller, D. The role of databases in support of computational chemistry calculations. *J. Comput. Chem.* **1996**, *17*, 1571–1586.

(55) Schuchardt, K. L.; Didier, B. T.; Elsethagen, T.; Sun, L.; Gurumoorthi, V.; Chase, J.; Li, J.; Windus, T. L. Basis Set Exchange: A Community Database for Computational Sciences. *J. Chem. Inf. Model.* **2007**, *47*, 1045–1052.

(56) Shahbazian, S.; Zahedi, M. Towards a complete basis set limit of Hartree-Fock method: correlation-consistent versus polarized-consistent basis sets. *Theor. Chem. Acc.* **2005**, *113*, 152–160.

(57) Boese, A. D.; Martin, J. M. L.; Handy, N. C. The role of the basis set: Assessing density functional theory. *J. Chem. Phys.* **2003**, *119*, 3005–3014.

(58) Schäfer, A.; Horn, H.; Ahlrichs, R. Fully optimized contracted Gaussian basis sets for atoms Li to Kr. *J. Chem. Phys.* **1992**, *97*, 2571–2577.

(59) Weigend, F.; Furche, F.; Ahlrichs, R. Gaussian basis sets of quadruple zeta valence quality for atoms H-Kr. *J. Chem. Phys.* **2003**, *119*, 12753–12762.

(60) Klopper, W.; Noga, J.; Koch, H.; Helgaker, T. Multiple basis sets in calculations of triples corrections in coupled-cluster theory. *Theor. Chem. Acc.* **1997**, *97*, 164–176.

(61) Karton, A.; Taylor, P. R.; Martin, J. M. L. Basis set convergence of post-CCSD contributions to molecular atomization energies. *J. Chem. Phys.* **2007**, *127*, 064104.

(62) Lee, T. J.; Taylor, P. R. A diagnostic for determining the quality of single-reference electron correlation methods. *Int. J. Quantum. Chem., Quantum. Chem. Symp.* **1989**, *23*, 199–207.

(63) Janssen, C. L.; Nielsen, I. M. B. New diagnostics for coupled-cluster and Møller-Plesset perturbation theory. *Chem. Phys. Lett.* **1998**, *290*, 423–430.

(64) Lévy-Leblond, J.-M. Nonrelativistic particles and wave equations. *Commun. Math. Phys.* **1967**, *6*, 286–311.

(65) Reiling, S.; Brickmann, J.; Schlenkrich, M.; Bopp, P. A. 1,2-Ethanediol: The Problem of Intramolecular Hydrogen Bonds. *J. Comput. Chem.* **1996**, *17*, 133–147.

(66) Jensen, F. The magnitude of intramolecular basis set superposition error. *Chem. Phys. Lett.* **1996**, *261*, 633–636.

(67) Senent, M. L.; Wilson, S. Intramolecular Basis Set Superposition Errors. *Int. J. Quantum Chem.* **2001**, *82*, 282–292.

(68) Boys, S. F.; Bernardi, F. The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Mol. Phys.* **1970**, *19*, 553–566.

(69) Dos Santos, H. F.; Rocha, W. R.; De Almeida, W. B. On the evaluation of thermal corrections to gas phase ab initio relative energies: implications to the conformational analysis study of cyclooctane. *Chem. Phys.* **2002**, *280*, 31–42.

CT800182E

JCTC Journal of Chemical Theory and Computation

Nonenzymatic Pathway of PUFA Oxidation. A First-Principles Study of the Reactions of OH Radical with 1,4-Pentadiene and Arachidonic Acid

Milan Szori,[‡,‡] Imre G. Csizmadia,[‡,§] and Bela Viskolcz*,[‡]

Department of Chemical Informatics, Faculty of Education, University of Szeged, Boldogasszony sgt. 6, 6725 Szeged, Hungary, Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Flemingovo náméstí 2, 16610 Prague 6, Czech Republic, and Department of Chemistry, University of Toronto, Toronto, Ontario, Canada M5S 3H6

Received April 14, 2008

Abstract: The oxidation of polyunsaturated hydrocarbons by •OH radical can play an important role in lipid oxidation of polyunsaturated fatty acids (PUFA) such as arachidonic acid (AA). As a prototype of this oxidation, the reaction of 1,4-pentadiene with the •OH radical is studied using the QCISD(T)/cc-pVTZ//BH&HLYP/6−31G(d) level of theory. One of the prereaction complexes is shown to be a springboard for the indirect bisallylic hydrogen abstraction (A_O), terminal ($T0_O$), and nonterminal •OH addition ($NT0_O$) reactions. The enthalpies of the transition states of the A_O, $T0_O$, and $NT0_O$ reactions are found to be lower than those of the reactants, so all these reactions are expected to be fast. The nonterminal adduct is found to be reactive *via* two low-lying consecutive reaction channels. The first channel is a five-membered ring closing ($NT1_O$). The second channel is bond scission, which results in an allyl radical and a vinyl alcohol ($NT2_O$). An analogous reaction pathway in which AA takes the place of 1,4-pentadiene was explored using the ONIOM(QCISD(T)/cc-pVTZ:BH&HLYP/6−31G(d))//BH&HLYP/6−31G(d) method. The results show that the formation of the five-membered ring (AA-$NT1_O$) is energetically favored. Our results demonstrate for the first time a possible, *ab initio*-based mechanism for the nonenzymatic biosynthesis of isoprostane-like structures from AA without the presence of molecular oxygen. Furthermore, the energetically low-lying bond scission channel may explain the observed formation of short fatty acids and dieneols (tautomers of unsaturated aldehydes).

1. Introduction

Fatty acids (FAs) are essential components of membrane phospholipids. The hydrocarbon chain of the FAs can be saturated as well as either monounsaturated (MUFA) or polyunsaturated (PUFA). An important subclass of PUFAs, called $\omega-6$ fatty acids, can be characterized by a double bond network which starts from the sixth carbon−carbon bond counting from the methyl carbon at the tail end (ω end) of the fatty acid.

PUFAs are integral structural components of membrane phospholipids, where they play an important role in maintaining the structural and functional characteristics of bilayer cell membranes within their homeostatic boundaries. Besides their physical effects on the membrane, PUFAs also contribute to regulatory function through eicosanoid production.[1] Eicosanoids are physiologically active compounds derived biosynthetically from 20-carbon fatty acids following their release from the membrane, through the specific action of phospholipase A2, and include prostaglandins, thromboxanes, and leukotrienes.[2]

Arachidonic acid (AA, 20:4n-6) is an $\omega-6$ fatty acid which is highly concentrated in brain tissue as an essential

* Corresponding author e-mail: viskolcz@jgypk.u-szeged.hu.

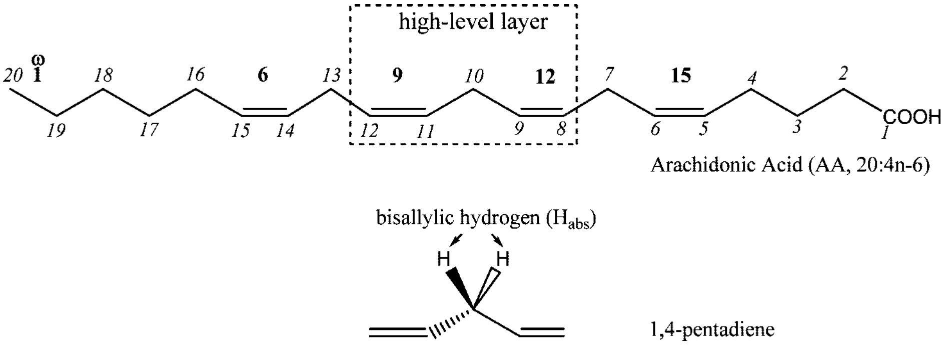
[‡] University of Szeged.

[‡] Academy of Sciences of the Czech Republic.

[§] University of Toronto.

10.1021/ct800127a CCC: $40.75 © 2008 American Chemical Society
Published on Web 08/02/2008

Scheme 1. Structure of Arachidonic Acid and 1,4-Pentadiene[a]



[a] The box with dotted lines shows the definition of the high-level layer in the ONIOM calculation.

component of membrane phospholipids.[3] Arachidonic acid is the most important prostaglandin precursor in humans,[4] and it is a component of the inositol phospholipids. AA plays a role in eicosanoid synthesis, such as the biosynthesis of 15-F_2-isoprostane. The conventional view of isoprostane synthesis proceeds *via* a pathway,[5] which begins with the formation of a radical by abstraction of the bisallylic hydrogen (H_{abs}) on C13 of AA (Scheme 1). Then, after several reaction steps, a five-membered ring is formed from C8−C12 carbons by means of molecular oxygen, producing isoprostanes.

There are several shortcomings to this view. The formation of isoprostanes has been detected in the nonenzymatic oxidation of the membrane *in vivo* and *in vitro*.[6,7] Furthermore, the conventional view cannot account for observation of unsaturated aldehydes in the oxidation of the membrane compounds.[8] However, mechanistic steps for the nonenzymatic pathway have not been established theoretically.

Our aim in this work is to gain a better understanding of the nonenzymatic pathway. From a reactivity point of view, 1,4-pentadiene (Scheme 1) is a good candidate for elucidating possible oxidation pathways of PUFAs. The reaction 1,4-pentadiene + •OH has other advantages, such as the moderate computational effort needed to perform the calculation and the absence of steric effects.

This paper is organized as follows. Section 2 describes the quantum chemical methods used for geometry optimizations and high-level energy calculation. Section 3 describes the prereaction complexes and the associated low-lying channels for the 1,4-pentadiene + •OH reaction, including transition states and products. Having established possible pathways for 1,4-pentadiene, similar calculations are carried out for arachidonic acid (AA) in place of 1,4-pentadiene.

2. Methods

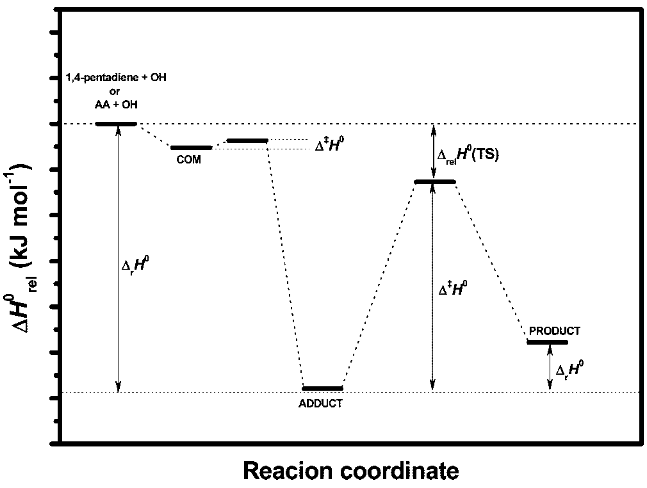
All calculations were performed by using the Gaussian03 program package.[9] In the case of reactions between alkenes and the hydroxyl radical, it has been shown previously[10] that the BH&HLYP functional gives reasonable geometries when used in combination with the 6−31G(d) split-valence basis set.[11] In every geometry optimization, 'Tight' convergence criteria were used. The nature of stationary points was checked by means of frequency calculations. Analytical second derivatives of the energy with respect to Cartesian coordinates were used for the determination of vibrational frequencies. Furthermore, additional and accurate single point calculations were carried out on the BH&HLYP/6−31G(d) geometries using the QCISD(T) method[12,13] along with the cc-pVTZ Dunning basis set.[14] Intrinsic reaction coordinate (IRC) calculation started from the A_O transition state was also computed using the BH&HLYP/6−31G(d) level of theory.

The structures in reactions of AA + •OH were also optimized at the BH&HLYP/6−31G(d) level of theory. Accurate single point energies were computed using the two-layer ONIOM technique,[15] employing the QCISD(T)/cc-pVTZ and the BH&HLYP/6−31G(d) level of theories for the high-level and low-level layers, respectively. The high-level layer contains C8−C12 and connected hydrogens, as illustrated in Scheme 1. These calculations are referred to as ONIOM(QCISD(T)/cc-pVTZ:BH&HLYP/6−31G(d))//BH&HLYP/6−31G(d).

Activation enthalpies ($\Delta^{\ddagger}H°$) are calculated as the difference of QCISD(T)/cc-pVTZ energies between the transition state (TS) structure and the van der Waals complex (COM), adding the difference of their corresponding enthalpies calculated at the BH&HLYP/6−31G(d) level of theory, and without any scale correction (Scheme 2). The electronic partition function of a species is assumed to be equal to its multiplicity. In the case of the consecutive steps ($NT1_O$ and $NT2_O$ as well as AA-$NT1_O$ and AA-$NT2_O$), the nonterminal adduct is regarded as the reactant, rather than the prereaction complex (COM). The relative enthalpies of the transition state ($\Delta H°_{rel}$(TS)) are related to the enthalpy level of the reactants (1,4-pentadiene and hydroxyl radical). The same nomenclature is used in the case of the AA + •OH reaction; the only difference is the ONIOM(QCISD(T)/cc-pVTZ:BH&HLYP/6−31G(d)) energy term instead of QCISD(T)/cc-pVTZ.

In the case of AA + •OH reactions, the effects of the surrounding lipid are also estimated though BH&HLYP/6−31G(d) single-point calculation using the CPCM model.[16] The solute cavity is built up by means of radii from the UFF force field (RADII = UFF). A value of 4.0 was employed for the dielectric constant (ϵ), to simulate the hydrophobic interior of a lipid membrane. All the remaining parameters

Scheme 2. Schematic Picture about the Definition of Different Thermodynamic Properties[a]



[a] Enthalpy, *H*, is the example, and it is also valid for entropy and Gibbs free energy.

for the surrounding lipid are the same as those for water. The contribution of the solvation to activation and reaction Gibbs free energies, $\Delta\Delta G^{\circ}_{solv}(X)$, are formulated as

$$\Delta\Delta G^{\circ}_{solv}(X) = \Delta G^{\circ}_{solv}(X) - \Delta G^{\circ}_{solv}(AA + \cdot OH)$$

3. Results and Discussion

Reactions of 1,4-Pentadiene and Hydroxyl Radical. It is well-known that in alkene + $\cdot$OH reactions,[10,17,18] the first step is the formation of a prereaction complex (COM). Depending on the orientation of the $\cdot$OH radical, one can distinguish two possible van der Waals complexes (COM_O and COM_I in Figure 1). In the outer complex (COM_O), the $\cdot$OH radical is further from the center of the mass of the complex. For the inner one (COM_I), the $\cdot$OH radical is closer to the center of mass of the van der Waals complex. Despite this structural difference, both complexes exhibit rather similar standard reaction enthalpies (Table 1). The outer complex, COM_O, is the initial structure for the terminal (T_O) and nonterminal additions (NT_O) as well as for the indirect hydrogen abstraction (A_O) reactions. In contrast to this, the indirect hydrogen abstraction cannot take place in COM_I, since the distance between the closest hydrogen and the oxygen is too large (3.876 Å) as can be seen in Figure 1. The corresponding distance in COM_O is only 2.826 Å, but the two van der Waals complexes are almost identical in the rest of the geometrical parameters. In the IRC calculation started from transition state of A_O, the $\cdot$OH radical approaches the double bond of the 1,4-pentadiene which shows that there is a pathway between A_O and COM_O.

The difference between the standard reaction entropies for the complex formation is 6.3 J mol^{-1} K^{-1}, which might be due to the interaction of the oxygen with the bisallylic hydrogen (H_{abs}). This difference has only a small contribution to the standard reaction Gibbs free energy (Table 1). Our calculations show that there is no significant energetic difference between the existing reaction channels started from the COM_O and the COM_I (the maximum deviation is smaller than 2.5 kJ mol^{-1}). Thermodynamic properties of all these reactions are listed in Table 1. Because the addition reactions from COM_O and COM_I differ in chirality and structural parameters, so they differ slightly in enthalpy. Due to this, the order of channels in the activation enthalpy can be changed. These effects are rather small.

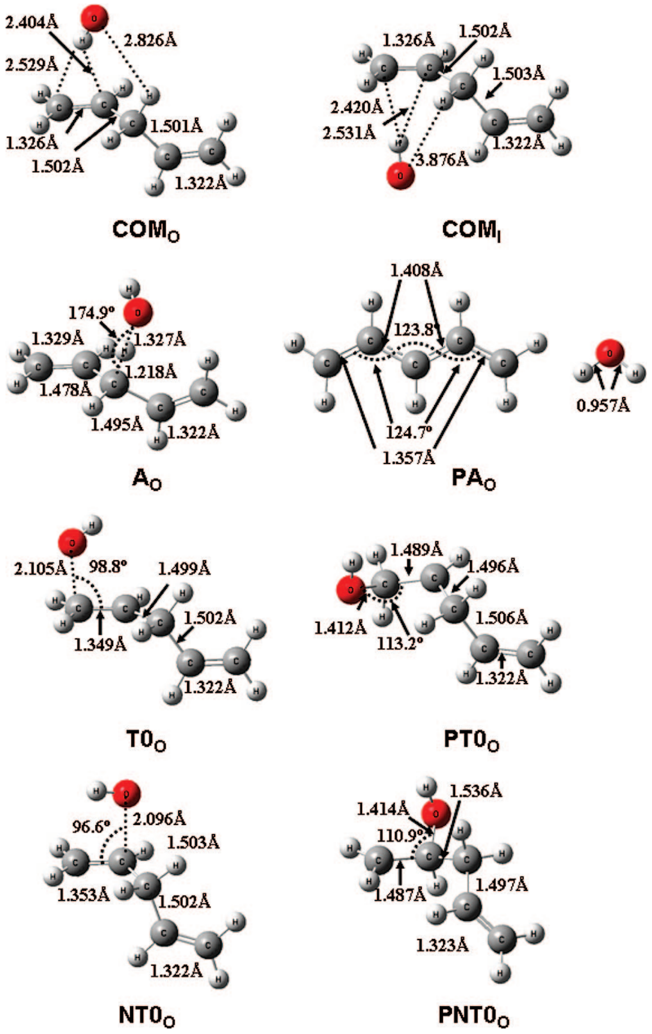


Figure 1. BH&HLYP/6−31G(d) optimized structures of the inner (COM_I) and outer (COM_O) prereaction complexes as well as minimum and transition state structures of the terminal addition ($T0_O$, $PT0_O$), indirect abstraction ($A_{O,}$ PA_O), and nonterminal addition ($NT0_O$, $PNT0_O$) channels in the 1,4-pentadiene + $\cdot$OH reaction system.

As Figure 2 shows, all of channels associated with COM_O have transition states that are below the entrance enthalpy level. The energetically most favored channel is the indirect hydrogen abstraction (A_O) reaction ($\Delta H^{\circ}_{rel} = -8.0$ kJ mol^{-1}, $\Delta^{\ddagger}H^{\circ} = 2.5$ kJ mol^{-1}). The products of this reaction are the resonance stabilized 1,4-pentadien-3-yl and a water molecule (PA_O). This reaction can also be considered as a prototype for bisallylic H-abstraction reactions, since the 1,4-pentadien-3-yl radical has five electrons delocalized on five carbon atoms,[19] which is a conjugated diallyl radical. The reaction is strongly exothermic (-167.9 kJ mol^{-1}). Its transition state structure (A_O) shows strongly reactant-like behavior, since the O−H_{abs} distance (1.327 Å) is significantly larger than the O−H distance (0.957 Å) in the water molecule. The

Table 1. Standard Reaction, Activation, and Relative Enthalpies ($\Delta_rH°$, $\Delta^{\ddagger}H°$, and $\Delta H°_{rel}$(TS) in kJ mol^{-1}) and Standard Reaction and Activation Entropies ($\Delta_rS°$ and $\Delta^{\ddagger}S°$ in J mol^{-1}K^{-1}) for the Reactions of 1,4-Pentadiene with the Hydroxyl Radical, Obtained from Calculations Performed at the QCISD(T)/cc-pVTZ//BH&HLYP/6-31G(d) Level of Theory

	$\Delta_rH°$ (kJ mol^{-1})	$\Delta_rS°$ (J mol^{-1} K^{-1})	$\Delta_rG°$ (kJ mol^{-1})	$\Delta H°_{rel}$(TS) (kJ mol^{-1})	$\Delta^{\ddagger}H°$ (kJ mol^{-1})	$\Delta^{\ddagger}S°$ (J mol^{-1} K^{-1})	$\Delta^{\ddagger}G°$ (kJ mol^{-1})
				Outer			
COM_O	−10.5	−99.7	19.2	-	-	-	-
A_O	−167.9	5.3	−169.5	−8.0	2.5	−22.8	9.3
$T0_O$	−110.4	−128.5	−72.1	−5.7	4.8	−22.7	11.6
$NT0_O$	−115.9	−136.1	−75.3	−7.3	3.2	−28.5	11.7
$NT1_O$	−78.4	−32.4	−68.7	−45.4	70.5	−32.6	80.2
$NT2_O$	20.3	157.3	−26.6	−25.5	90.4	0.4	90.3
				Inner			
COM_I	−11.8	−106.0	19.8	-	-	-	-
$T0_I$	−111.3	−131.3	−72.2	−9.2	2.6	−25.7	10.3
$NT0_I$	−118.2	−139.1	−76.8	−6.9	4.9	−26.2	12.7
$NT1_I$	−76.1	−29.5	−67.3	−47.9	70.3	−28.7	78.9
$NT2_I$	22.6	161.3	−25.4	−26.6	91.7	2.1	91.0

C−H$_{abs}$ bond being broken (1.218 Å) is larger by 0.128 Å than that in 1,4-pentadiene (1.090 Å). Due to the exothermicity of this reaction, one may expect a low probability of occurrence for the reverse reaction, an expectation that is reinforced by the fact that the forward reaction has a large negative standard Gibbs free energy (−169.5 kJ mol^{-1}, Table 1).

The reaction with the highest activation enthalpy is the terminal addition ($T0_O$). If one compares the enthalpy of $T0_O$ to that of COM_O, then the enthalpy barrier is 4.8 kJ mol^{-1}. In this case, the product is a stable 4-pentenyl-1-ol radical ($PT0_O$). As one can see from Figure 1, the oxygen of the •OH radical is still far (2.105 Å) from the terminal carbon in the transition state of the terminal addition ($T0_O$). In the case of the product, $PT0_O$, the C−O distance becomes shorter (1.412 Å), while the C═C bond extends to 1.489 Å, so it becomes more single bond like. The O−C−C angle increases from 98.8° to 113.2°.

The transition states for the possible consecutive reaction steps from the adducts $PT0_O$ and $PNT0_O$ were also characterized at the BH&HLYP/6−31G(d) level of theory (see the Supporting Information). In most cases, their relative enthalpies were found to be 20 kJ mol^{-1} higher than the entrance enthalpy. As it will be shown later on, these channels are energetically unfavored in contrast to the two corresponding channels of the nonterminal adduct. Due to this fact, contribution of these channels to the overall kinetics is expected to be negligible at room temperature, so they are not considered further.

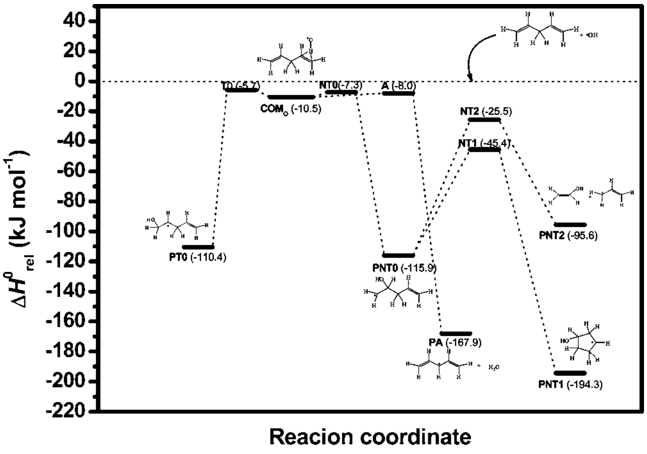


Figure 2. Standard enthalpy diagram for the energetically preferred reactions of 1,4-pentadiene with •OH. The values are obtained at the QCISD(T)/cc-pVTZ//BH&HLYP/6−31G(d) level of theory.

The third reaction channel studied is the nonterminal addition reaction ($NT0_O$). Its enthalpy is 3.2 kJ mol^{-1} relative to that of COM_O (Table 1). The geometrical parameters of the nonterminal transition state are quite similar to that of the terminal one (Figure 1). In this case, the oxygen of the hydroxyl radical is also found to be far from the carbon, 2.096 Å. The C−C−O angle is also close to perpendicular, 96.6°.

Although the geometry of the 4-pentene-2-ol-1-yl radical formed ($PNT0_O$) also shows similarities to $PT0_O$, it is more reactive, since there are two possible low-lying transition states. The energetically preferred one is the ring closing reaction, $NT1_O$, which gives the cyclopentanol-3-yl radical, $PNT1_O$. Although the enthalpy barrier of this reaction seems to be rather large, 70.5 kJ mol^{-1}, it is still below the enthalpy level of the entrance channel, −45.4 kJ mol^{-1}. It is also interesting to note that the ring closing reaction is also exothermic ($\Delta_rH°(NT1_O)$ = −78.4 kJ mol^{-1}). The enthalpy level of the cyclopentanol-3-yl radical is 26.4 kJ mol^{-1} lower than that of the product of the hydrogen abstraction, so the product radical is 194.3 kJ mol^{-1} more stable compared to the level of 1,4-pentadiene and hydroxyl radical. In its transition state (Figure 3), the distance between the two terminal carbons is 2.231 Å, which is about 0.7 Å larger than that found in the product (1.533 Å). The latter distance is in accord with a single carbon−carbon bond. The C−C•••C and C•••C−C angles also change significantly on going from the transition state to the product (Figure 3), the former extending from 87.8° to 104.2°, while the latter is 88.6° in the transition state structure and 103.4° in the product.

The other channel is the carbon−carbon bond scission, which gives vinyl alcohol as well as allyl radical as products ($PNT2_O$). Its transition state is product-like, as one can see from Figure 3. The broken bond distance is somewhat larger than 2.0 Å in the transition state, 2.096 Å. The enthalpy of

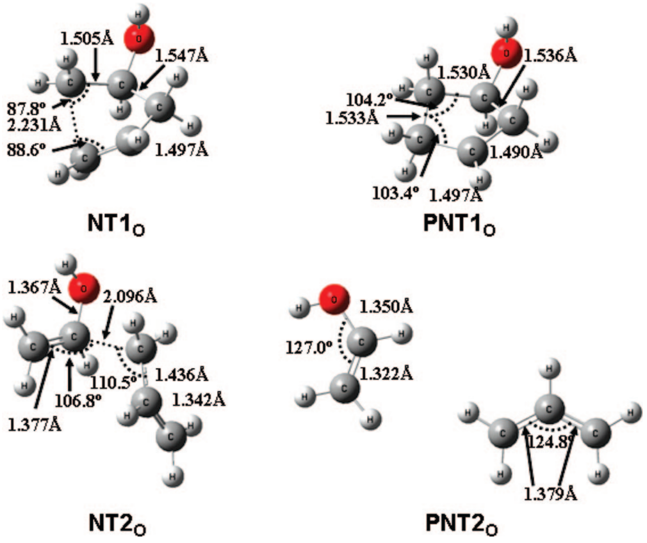


Figure 3. BH&HLYP/6−31G(d) optimized transition state structures for the five-membered ring formation reaction ($NT1_O$) and for the bond dissociation reaction ($NT2_O$) from the nonterminal addition adduct ($PNT0_O$).

$NT2_O$ is 90.4 kJ mol^{-1} relative to $PNT0_O$, and the reaction itself is slightly endothermic ($\Delta_r H°(NT2_O) = 22.6$ kJ mol^{-1}). Surprisingly, the standard activation entropy is close to zero (0.4 J mol^{-1} K^{-1}), and the reverse reaction has also a relatively large enthalpy barrier (70.1 kJ mol^{-1}). This value indicates that the bimolecular reverse reaction is not favored energetically. Thermodynamically, $PNT2_O$ is favored over $PNT0_O$ at room temperature.

Interestingly, 1,4-pentadiene is not only important as a model of PUFAs but also is found in mineral and vegetable oils. Consequently, its oxidation by hydroxyl radical also has a great impact on combustion chemistry.[20,21] Furthermore, the polyunsaturated hydrocarbons are also known to be present in the atmosphere as anthropogenic and biogenic volatile organic compounds (VOC).[22,23]

Reactions of Arachidonic Acid (AA) with Hydroxyl Radical. Given its relevance, the oxidation of arachidonic acid (AA) is frequently assayed using a wide variety of experimental setups, including HPLC/MS/MS.[6,7] Although these experiments are state-of-the-art, only stable products can be detected. Since abstraction of hydrogen by the •OOH radical cannot be fast at the bisallylic position,[10] the •OH radical is considered as a radical source in the first step of lipid peroxidation.[5] This prompted us to study possible initial steps for the AA oxidation by the •OH radical.

For several $\omega-3$ PUFA species, different conformations were studied. The extended conformer was found to be the global minimum in vacuum.[24] Based on this result, the extended conformer of AA was used as an initial structure in our work. The reactions studied can be found in Figure 4, using the analogy of the 1,4-pentadiene + •OH reaction involving hydrogen abstraction (AA-A_O) and nonterminal-like addition (AA-$NT0_O$) and its consecutive steps (AA-$NT1_O$ and AA-$NT2_O$). All of these channels were found, and the optimized structures of their corresponding critical points are shown in Figure 5. There are four double bonds in AA, and each may form a van der Waals complex (the •OH radical can approach the AA at the $\omega-6$, $\omega-9$, $\omega-12$, and $\omega-15$ positions, using the notation of Scheme 1). We only consider the AA + •OH reactions *via* the $\omega-9$ van der Waals complex in this work (Figure 4), since our aim is to show the mechanism of formation of isoprostane-like species. There is no doubt that reactions involving other types of prereaction complexes ($\omega-6$ or $\omega-12$ or $\omega-15$) can also occur, since the bond energies of bisallylic hydrogens do not depend strongly on their position in the chain.[25] In these cases, only the positions of the five-membered rings in AA-$PNT1_O$ as well as the sizes of the carbon chains in the products of AA-$PNT2_O$ would vary.

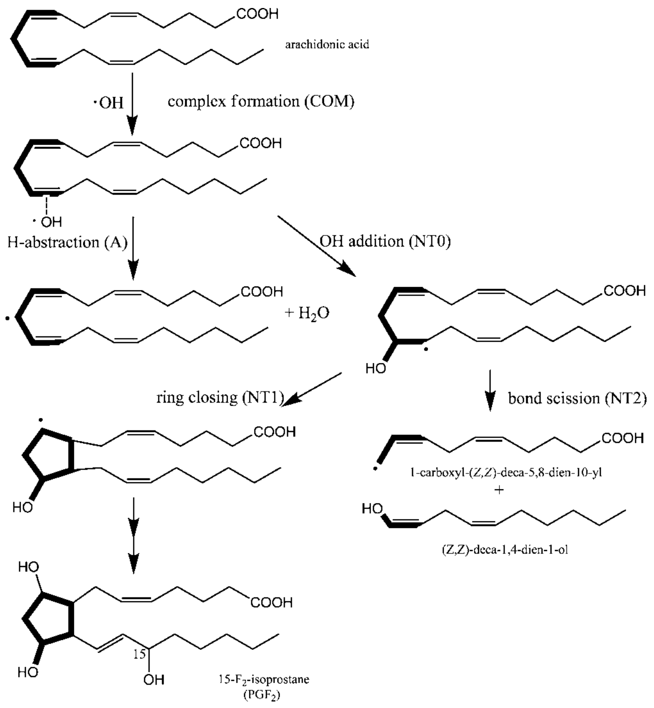


Figure 4. Possible nonenzymatic pathways for oxidation of arachidonic acid (AA) as a lipid-oxidation example. The pathways are suggested based on the radical reaction of 1,4-pentadiene with the hydroxyl radical. The pattern of 1,4-pentadiene is indicated by thick lines.

Comparison of Figure 5 with Figure 1 and Figure 3 shows generally a good structural agreement between the 1,4-pentadiene and the AA reactions. Two exceptions stand out: the different is a little larger in comparisons of AA-COM_O vs COM_O as well as AA-A_O vs A_O, respectively. The •OH and C=C double bond are somewhat closer in the AA-COM_O. This could be the consequence of the steric effects in the case of weakly bonded structures. The transition state of the abstraction reaction (AA-A_O in Figure 5) seems to be of a more earlier type compared to that of 1,4-pentadiene and •OH (A_O in Figure 1), since the O•••H_{abs} is found to be 1.405 Å which is 1.327 Å in the case of the 1,4-pentadiene reaction. Furthermore, the C•••H_{abs} distance also becomes shorter by 0.029 Å.

If one compares the values in Table 1 to those in Table 2, the difference in enthalpy between 1,4-pentadiene and corresponding AA reactions are in general less than 11 kJ mol^{-1}. The single exception is the reaction enthalpy of the ring closing reaction, $NT1_O$, 27.1 kJ mol^{-1}. This difference

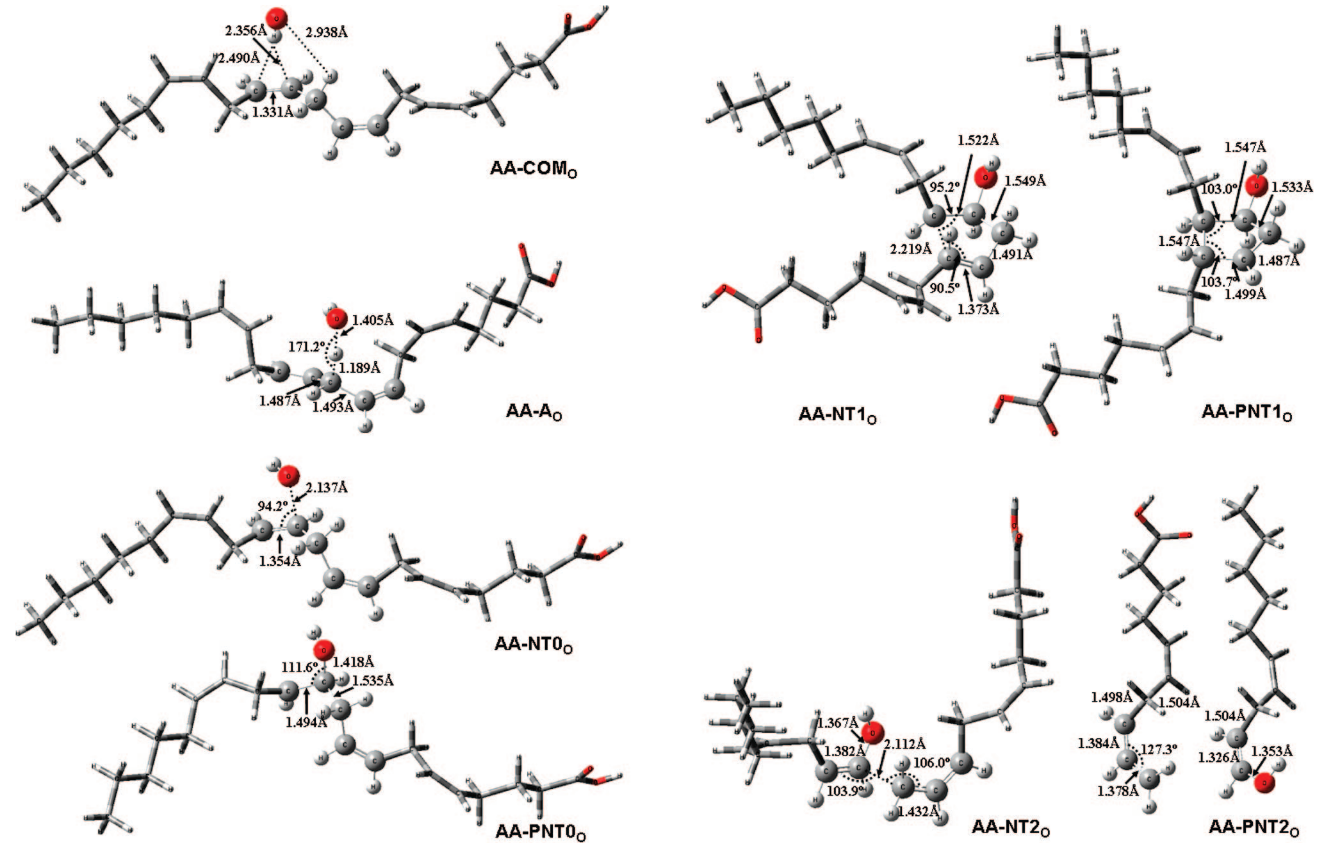


Figure 5. Transition state and minimum structures for the reactions of arachidonic acid (AA) with •OH radical, obtained at the BH&HLYP/6−31G(d) level of theory.

Table 2. Standard Reaction, Activation, and Relative Enthalpies ($\Delta_r H^\circ$, $\Delta^\ddagger H^\circ$, and ΔH°_{rel}(TS) in kJ mol^{-1}) and Standard Reaction and Activation Entropies ($\Delta_r S^\circ$ and $\Delta^\ddagger S^\circ$ in J mol^{-1} K^{-1}) for the Reactions of Arachidonic Acid (AA) with the Hydroxyl Radical, Obtained from Calculations at the ONIOM(QCISD(T)/cc-pVTZ:BH&HLYP/6-31G(d))//BH&HLYP/6-31G(d) Level of Theory[a]

	$\Delta_r H^\circ$ (kJ mol^{-1})	$\Delta_r S^\circ$ (J mol^{-1} K^{-1})	$\Delta_r G^\circ$ (kJ mol^{-1})	ΔH°_{rel}(TS) (kJ mol^{-1})	$\Delta^\ddagger H^0$ (kJ mol^{-1})	$\Delta^\ddagger S^0$ (J mol^{-1} K^{-1})	$\Delta^\ddagger G^0$ (kJ mol^{-1})
				Outer			
AA-COM$_O$	−13.1	−115.8	21.4 (4.7)	-	-	-	-
AA-A$_O$	−173.6	7.5	−175.9 (−10.8)	−10.0	3.1	−36.4	13.9 (−3.4)
AA-NT0$_O$	−123.2	−146.8	−79.4 (0.3)	−12.4	0.7	−16.8	0.7 (−0.1)
AA-NT1$_O$	−51.3	−46.5	−37.4 (−6.3)	−56.0	67.2	−39.5	78.9 (10.8)
AA-NT2$_O$	26.1	169.8	−24.5 (−3.9)	−29.4	93.8	3.7	92.7 (0.8)

[a] Values in parentheses are the contributions of the bulk solvation (lipid) to Gibbs free energy ($\Delta\Delta G^\circ_{solv}$) calculated by CPCM-BH&HLYP/6-31G(d) single point calculation (ϵ = 4.0).

may be due to a steric constraint caused by the ring formed in the reaction.

Furthermore, the AA-A$_O$ and AA-NT0$_O$ activation enthalpies are slightly negative (−1.5 and −3.0 kJ mol^{-1}, respectively). Compared to the activation enthalpies of 1,4-pentadiene and AA reaction systems, the difference is found to be as small as 0.6 kJ mol^{-1} for A$_O$, −2.5 kJ mol^{-1} for NT0$_O$, −3.3 kJ mol^{-1} for NT1$_O$, and 3.4 kJ mol^{-1} for NT2$_O$.

The most important result to emphasize is that the transition states for the AA + •OH reaction system are ring closing and bond scission, and these states lie energetically below the entrance enthalpy level (AA + •OH). The product of the ring closing reaction is AA-PNT1$_O$. The bond scission reaction produces a shortened fatty acid, 1-carboxyl-(Z,Z)-deca-5,8-diene-10-yl radical, and an eneol, the (Z,Z)-deca-1,4-diene-1-ol molecule (Figure 4 and Figure 5). The (Z,Z)-deca-1,4-diene-1-ol might convert to its tautomer, Z-4-ene-decanal. However, subsequent reactions become far too complex to be characterized since the number of possible reactions increases with the number of components.

The effects of the surrounding lipids are estimated using the CPCM model (Table 2) for the reaction system AA + •OH. The largest contribution of solvation to reaction Gibbs free energy is found in the case of the hydrogen abstraction ($\Delta\Delta_r G^\circ_{solv}$(AA-A$_O$) = −10.8 kJ mol^{-1}). All the remaining $\Delta\Delta_r G^\circ_{solv}$ are moderate with absolute values less than 7 kJ mol^{-1}. The largest contribution of solvation to activation Gibbs free energy (10.8 kJ mol^{-1}) is noted for the ring closing reaction (AA-NT1$_O$). For the other transition states, the solvation has no significant influence on the free energy profile.

Table 3. Standard Gibbs Free Energy for the Formation of the Prereaction Complex, $\Delta G^\circ(\mathrm{COM_O})$ and for the Indirect Hydrogen Abstraction Reaction ($\mathrm{A_O}$), $\Delta_r G^\circ$, As Well As the Activation Gibbs Free Energy of the Indirect Hydrogen Abstraction Reaction, $\Delta^\ddagger G^\circ$ (in kJ mol^{-1})[a]

systems		monoallylic	bisallylic
1,4-pentadiene + ·OH	$\Delta G^\circ(\mathrm{COM_O})$	-	19.2
A	$\Delta_r G^\circ(\mathrm{A_O})$	-	−169.5
this work	$\Delta^\ddagger G^\circ(\mathrm{A_O})$	-	9.3
(2*Z*,5*Z*)-heptadiene + ·OH	$\Delta G^\circ(\mathrm{COM_O})$	17.9	18.4
B	$\Delta_r G^\circ(\mathrm{A_O})$	−134.2	−182.2
ref 17	$\Delta^\ddagger G^\circ(\mathrm{A_O})$	29.3	5.7
(3*Z*,6*Z*)-nonadiene + ·OH	$\Delta G^\circ(\mathrm{COM_O})$	24.3	21.3
C	$\Delta_r G^\circ(\mathrm{A_O})$	−152.7	−169.0
ref 26	$\Delta^\ddagger G^\circ(\mathrm{A_O})$	30.5	17.6
arachidonic acid + ·OH	$\Delta G^\circ(\mathrm{COM_O})$	-	21.4
D	$\Delta_r G^\circ(\mathrm{A_O})$	-	−175.9
this work	$\Delta^\ddagger G^\circ(\mathrm{A_O})$	-	13.9

[a] A: QCISD(T)/cc-pVTZ//BH&HLYP/6-31G(d) level of theory. B: G3MP2//BH&HLYP/6-31G(d) level of theory. C: MPWB1K/MG3S//MPWB1K/6-31+G(d,p) level of theory. D: ONIOM(QCISD(T)/cc-pVTZ:BH&HLYP/6-31G(d))//BH&HLYP/6-31G(d) level of theory.

Initial steps of lipid peroxidation have been well studied with density functional methodology, both in the nonenzymatic process[26] as well as in the iron center of the (soybean) lipoxygenase enzyme.[27,28] Besides hydrogen abstraction, the addition of molecular oxygen is also included in the nonenzymatic study. These results, together with our previously published ones, are collected in Table 3. As one can see from this table, the Gibbs free energy of complex formation is in the range of 18.4 to 21.4 kJ mol^{-1}, depending on the methods used and the system studied; this difference is around the chemical accuracy. In contrast to this, the difference in the reaction Gibbs free energy of the bisallylic radical formation is about −169.0 kJ mol^{-1} for the 1,4-pentadiene + ·OH and (3Z,6Z)-nonadiene + ·OH reactions. The latter system was calculated at the MPWB1K//MG3S//MPWB1K/6−31+G(d,p) level of theory, and it was used by Tejero et al.[26] as a model for the lipid peroxidation of AA. The reaction Gibbs free energy of (2Z,5Z)-heptadiene + ·OH (calculated at the G3MP2//BH&HLYP/6−31G(d) level of theory) and AA + ·OH are also almost identical. The difference between the two groups is only 13 kJ mol^{-1}, which can be explained by the sum of several errors such as the difference in the computation methods. Surprisingly, the energetics of monoallylic and bisallylic hydrogen abstractions differ only by 16.3 kJ mol^{-1} in the case of the (3Z,6Z)-nonadiene + ·OH reaction system. That difference is 48 kJ mol^{-1} in the case of the (2Z,5Z)-heptadiene + ·OH reaction calculated at the G3MP2//BH&HLYP/6−31G(d) level of theory.[17]

Borowski et al.[27] published their theoretical work on the enzymatic process including the abstraction reaction and the O_2 addition. They studied the reaction between (Z,Z)-hepta-2,5-diene and the active site of the soybean enzyme (SLO-1). The initial structure of the active site is characterized by X-ray measurements. In this case, geometries were characterized using the B3LYP density functional and the LanL2DZ basis set with an effective core potential (ECP) for the Fe atom and the D95 basis set for H, C, N, and O atoms. The B3LYP/6−311+G(d,p) level of theory was used for single point calculations. The first catalytic step consisted of a hydrogen atom transfer from the hydrocarbon to the hydroxide group bound to the ferric ion. This process proceeds *via* an early transition state with the activation energy amounting to 50.6 kJ mol^{-1}, and the reaction energy is found to be −52.7 kJ mol^{-1}. This activation barrier seems to be quite high compared to our results of nonenzymatic hydrogen abstraction (with pseudo negative activation barrier). On the one hand, this result might be due to the inaccuracy of the available computational level for iron containing species. On the other hand, increased activation barrier in the enzymatic process can be the price paid for site-selective oxidation.

Tejero et al. also studied the (Z,Z)-hepta-2,5-diene + ·OH reaction[28] using a similar model for the enzymatic surroundings to that used by Borowski. They reported that the standard Gibbs free energy barrier for the hydrogen abstraction was as high as 87.0 kJ mol^{-1}, obtained using the B3LYP/6−311+G(d,p)//B3LYP/LanL2DZ level of theory. However, the differences between Borowski's and Tejero's results might arise mainly from conformational differences or because the extended model involving iron could be calculated only at lower accuracy.

The evolution of living organisms in an oxidizing atmosphere has resulted in a complex array of antioxidation mechanisms within cells to protect critical biomolecules from oxidative modifications. Because lipids are often the initial barrier to the free diffusion of reactive oxygen species into the cell, they themselves become targets of nonenzymatic oxidation reactions.[29] These nonenzymatic processes result in a wide variety of products having diverse biological functions, such aldehydes, shorter fatty acids, and isoprostanoids. While the first two groups are toxic, isoprostanoids with appropriate chirality (and within a certain concentration range) are essential for living cells. In addition, our calculations suggest that nonenzymatic formation of isoprostanoids is energetically favored. Based on these facts, we might suspect that in the course of biomolecular evolution, these nonenzymatic processes predated the enzymatic ones. Specific enzymes might have evolved for the purpose of controlling selectivity (the required products and their appropriate stereochemistry).

4. Conclusion

Employing the 1,4-pentadiene + ·OH reaction system, the energetically preferred oxidation pathways are studied for the PUFA using first-principles methods. We find that the terminal and nonterminal additions and the indirect hydrogen abstraction reaction have pseudonegative activation enthalpies due to the prereaction complex. The H-abstraction is found to be the most exothermic reaction among those studied (−167.9 kJ mol^{-1}).

The nonterminal adduct (the 4-pentene-2-ol-1-yl radical) is able to react forward to produce cyclopentanol-3-yl radical as product. The enthalpy level of the cyclopentanol-3-yl radical is 26.4 kJ mol^{-1} lower than that of the product of the hydrogen abstraction. The activation enthalpy of this ring closing reaction is significantly smaller than the exothermicity of the nonterminal addition. Consequently, this reaction is expected to be fast. Furthermore, there is also a possible channel for bond scission, which gives allyl radical and vinyl alcohol as products. Although this reaction is slightly endothermic, its activation enthalpy is still below the entrance enthalpy level of the 1,4-pentadiene + •OH reaction by 25.5 kJ mol^{-1}.

Since the PUFA have very often a −HC=CH−CH$_2$−CH=CH− moiety, one of them, arachidonic acid (AA), is studied as an analog to the reaction between 1,4-pentadiene and the hydroxyl radical. The thermodynamic properties obtained using the ONIOM technique for AA and QCISD(T)/cc-pVTZ//BH&HLYP/6−31G(d) for 1,4-pentadiene are found to be similar. Nonenzymatic ring closing and bond scission can also be energetically favored since the energetics of the transition states are still below the entrance enthalpy level (AA + •OH). As far as we know, our results demonstrate for the first time a possible, *ab initio*-based mechanism for the nonenzymatic biosynthesis of isoprostane-like structures from AA without the presence of molecular oxygen.

It is believed that the nonenzymatic ring closing and bond scission can also occur in the case of other PUFAs, such as docosahexaenoic acid (DHA).

Although these nonenzymatic radical reactions are energetically favored and they can occur in biological systems as spontaneous and fast processes, they are not selective. Specific enzymes might be responsible mainly for controlling the required products and their appropriate stereochemistry.

Acknowledgment. The authors thank Máté Labádi for excellent technical support. The authors are also grateful to Steven Neshyba and Claudio Morgado for their excellent suggestions.

Supporting Information Available: Reaction scheme for all possible indirect reactions of 1,4-pentadiene and •OH (indicated values are relative energies with zero-point correction related to the level of 1,4-pentadiene + •OH obtained by the BH&HLYP/6−31G(d) level of theory). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Sprecher, H. *Prog. Lipid Res.* **1986**, *25*, 19.

(2) Morrow, J. D.; Awad, J. A.; Wu, A.; Zackert, W. E.; Daniel, V. C.; Roberts, L. J. *J. Biol. Chem.* **1996**, *271*, 23185.

(3) Wainwright, P. E. *Neurosci. Biobehav. Rev.* **1992**, *16*, 193.

(4) Voet, D.; Voet J. *Biochemistry*; John Wiley & Sons Inc.: New York, NY, 1990; p 658.

(5) McMurry, J. E.; Begley, T. P. *The Organic Chemistry of Biological Pathways*; Roberts and Company Publisher: Englewood, CO, 2005; pp 364−368.

(6) Davis, T. A.; Gao, L.; Yin, H.; Morrow, J. D.; Porter, N. A. *J. Am. Chem. Soc.* **2006**, *128*, 14897.

(7) Yin, H.; Gao, L.; Tai, H. H.; Murphey, L. J.; Porter, N. A.; Morrow, J. D. *J. Biol. Chem.* **2007**, *282*, 329.

(8) Yoshino, K.; Sano, M.; Fujita, M.; Tomita, I. *Chem. Pharm. Bull.* **1991**, *39*, 1788.

(9) 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. *GAUSSIAN03, (Revision C.02)*; Gaussian, Inc.: Wallingford, CT, 2004.

(10) Szori, M.; Fittschen, C.; Csizmadia, I. G.; Viskolcz, B. *J. Chem. Theory Comput.* **2006**, *2*, 1575.

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

(12) Gauss, J.; Cremer, D. *Chem. Phys. Lett.* **1988**, *150*, 280.

(13) Salter, E. A.; Trucks, G. W.; Bartlett, R. J. *J. Chem. Phys.* **1989**, *90*, 1752.

(14) Kendall, R. A., Jr.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796.

(15) Dapprich, S.; Komaromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. J. *J. Mol. Struct. (THEOCHEM)* **1999**, *461*, 1.

(16) Barone, V.; Cossi, M. *J. Phys. Chem. A* **1998**, *102*, 1995.

(17) Szori, M.; Abou-Abdo, T.; Fittschen, C.; Csizmadia, I. G.; Viskolcz, B. *Phys. Chem. Chem. Phys.* **2007**, *9*, 1931.

(18) Sosa, C.; Schlegel, H. B. *J. Am. Chem. Soc.* **1987**, *109*, 4193.

(19) Szori, M.; Viskolcz, B. *J. Mol. Struct. (THEOCHEM)* **2003**, *666*, 153.

(20) Battin-Leclerc, F. *Phys. Chem. Chem. Phys.* **2002**, *4*, 2072.

(21) Skjoth-Rasmussen, M. S.; Glarborg, P.; Ostberg, M.; Johannessen, J. T.; Livbjerg, H.; Jensen, A. D.; Christensen, T. S. *Combust. Flame* **2004**, *136*, 91.

(22) Zielinska, B.; Sagebiel, J. C.; Harshfield, G.; Gertler, A. W.; Pierson, W. R. *Atmos. Environ.* **1996**, *30*, 2269.

(23) Reducing Emissions. 2005. Canadian Chemical Producers' Association Web Site. http://www.ccpa.ca/files/Library/Reports/NERM14_2005/RE14_Report07EN.pdf (accessed June, 4, 2008).

(24) Law, J. M. S.; Szori, M.; Izsak, R.; Penke, B.; Csizmadia, I. G.; Viskolcz, B. *J. Phys. Chem. A* **2006**, *110*, 6100.

(25) Kitaguchi, H.; Ohkubo, K.; Ogo, S.; Fukuzumi, S. *Chem. Commun.* **2006**, *9*, 979.

(26) Tejero, I.; González-Lafont, A.; Lluch, J. M.; Eriksson, L. A. *J. Phys. Chem. B* **2007**, *111*, 5684.

(27) Borowski, T.; Brocławik, E. *J. Phys. Chem. B,* **2003**, *107*, 4639.

(28) Tejero, I.; Eriksson, L. A.; González-Lafont, A.; Marquet, J.; Lluch, J. M. *J. Phys. Chem. B* **2004**, *108*, 13831.

(29) Smith, W. L.; Murphy, R. C. *J. Bio. Chem.* **2008**, *283*, 15513.

CT800127A

JCTC Journal of Chemical Theory and Computation

Accurate Molecular Polarizabilities Based on Continuum Electrostatics

Jean-François Truchon,[†,‡] Anthony Nicholls,[§] Radu I. Iftimie,[†] Benoît Roux,[||] and Christopher I. Bayly*,[‡]

Département de chimie, Université de Montréal, C.P. 6128 Succursale centre-ville, Montréal, Québec, Canada H3C 3J7, Merck Frosst Canada Ltd., 16711 TransCanada Highway, Kirkland, Québec, Canada H9H 3L1, OpenEye Scientific Software, Inc., Santa Fe, New Mexico 87508, and Institute of Molecular Pediatric Sciences, Gordon Center for Integrative Science, University of Chicago, Illinois 929 East 57th Street, Chicago, Illinois 60637

Received April 7, 2008

Abstract: A novel approach for representing the intramolecular polarizability as a continuum dielectric is introduced to account for molecular electronic polarization. It is shown, using a finite-difference solution to the Poisson equation, that the electronic polarization from internal continuum (EPIC) model yields accurate gas-phase molecular polarizability tensors for a test set of 98 challenging molecules composed of heteroaromatics, alkanes, and diatomics. The electronic polarization originates from a high intramolecular dielectric that produces polarizabilities consistent with B3LYP/aug-cc-pVTZ and experimental values when surrounded by vacuum dielectric. In contrast to other approaches to model electronic polarization, this simple model avoids the polarizability catastrophe and accurately calculates molecular anisotropy with the use of very few fitted parameters and without resorting to auxiliary sites or anisotropic atomic centers. On average, the unsigned error in the average polarizability and anisotropy compared to B3LYP are 2% and 5%, respectively. The correlation between the polarizability components from B3LYP and this approach lead to a R^2 of 0.990 and a slope of 0.999. Even the F_2 anisotropy, shown to be a difficult case for existing polarizability models, can be reproduced within 2% error. In addition to providing new parameters for a rapid method directly applicable to the calculation of polarizabilities, this work extends the widely used Poisson equation to areas where accurate molecular polarizabilities matter.

1. Introduction

The linear response of the electronic charge distribution of a molecule to an external electric field, the polarizability, is at the origin of many chemical phenomena such as electron scattering,[1] circular dichroism,[2] optics,[3] Raman scattering,[4] softness and hardness,[5] electronegativity,[6] and so forth. In atomistic simulations, polarizability is believed to play an important and unique role in intermolecular interactions of heterogeneous media such as ions passing through ion channels in cell membranes,[7] in the study of interfaces,[8] and in protein–ligand binding.[9]

Polarizability is considered to be a difficult and important problem from a theoretical point of view. Much effort has been invested in the calculation of molecular polarizability at different levels of approximation. At the most fundamental level, electronic polarization is described by quantum mechanics (QM) electronic structure theory such as extended basis set density functional theory (DFT) and ab initio molecular orbital theory. However, the extent of the com-

* Corresponding author phone: (514) 428-3403; fax: (514) 428-4930; e-mail: christopher_bayly@merck.com.

[†] Université de Montréal.

[‡] Merck Frosst Canada Ltd.

[§] OpenEye Scientific Software, Inc.

[||] University of Chicago.

10.1021/ct800123c CCC: $40.75 © 2008 American Chemical Society
Published on Web 08/13/2008

putational resources required is an impediment to the wide application of these methods on large molecular sets or on large molecular systems such as drug-like molecules.[10] To circumvent these limitations, empirical physical models based on classical mechanics have been parametrized to fit experimental or quantum mechanical polarizabilities.

In this article, we explore a new empirical physical model to account for electronic polarizability in molecules. The electronic polarization from internal continuum (EPIC) model uses a dielectric constant and atomic radii to define the electronic volume of a molecule. The molecular polarizability tensor is calculated by solving the Poisson equation (PE) with a finite difference algorithm. The concept that a dielectric continuum can account for solute polarizability has been examined previously. For example, Sharp et al.[11] showed that condensed phase induced molecular dipole moments are accounted for with the continuum solvent approach and that it leads to accurate electrostatic free energy of solvation. More recently Tan and Luo[12] have attempted to find an optimal inner dielectric value that reproduces condensed phase dipole moments in different continuum solvents. In spite of these efforts, we found that none of these models can account correctly for molecular polarizability. Here, the concept is explored with the objective of producing a high accuracy polarizable electrostatic model. Therefore, we focus on the optimization of atomic radii and inner dielectrics to reproduce the B3LYP/aug-cc-pVTZ polarizability tensor.

In this preliminary work, we seek to establish the soundness and accuracy of the EPIC model in the calculation of the molecular polarizability tensor on three classes of molecules: homonuclear diatomics, heteroaromatics, and alkanes. These molecular classes required special attention with previous polarizable models as a result of their high anisotropy.[13−15] Overall, 53 different molecules are used to fit our model and 45 molecules to validate the results. These specific questions are addressed: Can the EPIC model accurately calculate the average polarizability? If so, can it further account for the anisotropy and the orientation of the polarizability components? How few parameters are needed to account for highly anisotropic molecules, and how does this compare to other polarizable models? How transferable are the parameters obtained with this model? Is the model able to account for conformational dependency? In answering these questions, we obtained a fast and validated method with optimized parameters to accurately calculate the molecular polarizability tensor for a large variety of heteroaromatics not previously considered.

The remainder of the article is organized as follows. In the next section, we briefly review the most successful existing polarizable approaches, focusing on aspects relevant to this study. Then we introduce the dielectric polarizable method with a polarizable sphere analytical model. A methodology section in which we outline the computational details follows. The molecular polarizability results are then reported. This is followed by a discussion and conclusion.

2. Existing Empirical Polarizable Models

2.1. Point Inducible Dipole. The point inducible dipole (PID) model was first outlined by Silberstein in 1902.[16] This model has been extensively used to calculate molecular polarizability[14,15,17−22] and to account for many-body effects in condensed phase simulations.[23−25] Typically, in the PID model, an atom is a polarizable site where the electric field direction and strength together with the atomic polarizability define the induced atomic dipole moment. Since the electric field at an atomic position is in part due to other atoms' induced dipoles, the set of equations must be solved iteratively (or through a matrix inversion). In 1972, Applequist[19] showed that the PID can accurately reproduce average molecular polarizability of a diverse set of molecules but also that the mathematical formulation of the PID can lead to a polarizability catastrophe. Briefly, when two polarizable atoms are close to each other, the solution to the mathematical equations involved is either undetermined (with the matrix inversion technique) or the neighboring dipole moments cooperatively increase to infinity. To circumvent this problem, Thole[14,22] modified the dipole field tensor with a damping function, which depends on a lengthscale parameter meant to represent the spatial extent of the polarized electronic clouds; his proposed exponential modification is still important and remains in use.[13,14,26]

2.2. Drude Oscillators. The Drude oscillator (DO) represents electronic polarization by introducing a massless charged particle attached to each polarizable atom by a harmonic spring.[27] When the Drude charge is large and tightly bound to its atom, the induced dipole essentially behaves like a PID. The DO model is attractive because it preserves the simple charge−charge radial Coulomb electrostatic term already present and it can be used in molecular dynamics simulation packages without extensive modifications. The DO model has not yet been extensively parametrized to reproduce molecular polarizability tensors, but recent results suggest that it could perform as well as PID methods. Finally, the DO model also requires a damping function to avoid the polarizability catastrophe.[26]

2.3. Fluctuating Charges. A third class of empirical model, called fluctuating charge (FQ), was first published in a study by Gasteiger and Marsili[28] in 1978 to rapidly estimate atomic charges. Subsequently, FQ was adapted to reproduce molecular polarizability and applied in molecular dynamic simulations.[29,30] It is based on the concept that partial atomic charges can flow through chemical bonds from one atomic center to another based on the local electrostatic environment surrounding each atom. The equilibrium point is reached when the defined atomic electronegativities are equal. The FQ model, like the DO, has mainly been used in condensed phase simulations and not specifically parametrized to reproduce molecular polarizabilities. A major problem with FQ is the calculation of directional polarizabilities (eigenvalues of the polarizability tensor). For planar or linear chemical moieties (ketones, aromatics, alkane chains, etc.) the induced dipole can only have a component in the plane of the ring or in line with the chain. For instance, the out-of-plane polarizability of benzene can only be correctly calculated if out-of-plane auxiliary sites are built. For alkane chains, though, there is no simple solution.[31] For this reason, the ability of the FQ model to accurately represent complex molecular polarizabilities is clearly limited.

2.4. Limitations with the PID Related Methods. The PID and the related models have been parametrized and show an average error on the average polarizability around 5%. However, errors in the anisotropy are often around 20% or higher.[15,20] Diatomic molecules are not handled correctly by any of these methods, leading to errors of 82% in the anisotropy for F_2, for example.[13,14] Heteroaromatics, which are abundant moieties in drugs, are often poorly described by PID methods. This limitation is due to the source of anisotropy in the PID model, that is, the interatomic dipole interaction located at static atom positions. It is nevertheless possible to improve these models. For example, using full atomic polarizability tensors instead of isotropic polarizabilities has reduced the errors in polarizability components from 20% to 7%.[20,21] In the case of the DO model, acetamide polarizabilities have been corrected by the addition of atom-type-dependent damping parameters and anisotropic harmonic springs.[32] In these cases, the improvement required a significant amount of additional parameters which brings an additional level of difficulty in their generalization. As illustrated below, our model seems to address most of these complications without additional parameters and complexity.

3. Dielectric Polarizability Model

The mathematical model that we explore in this article is based on simple concepts that have proved extremely useful in chemistry.[33−38] We propose a specific usage that we clarify and describe in this section.

3.1. Model. Traditionally in Poisson−Boltzmann (PB) continuum solvent calculations, the solute is described as a region of low dielectric containing a set of distributed point charges; the polar continuum solvent (usually water) is described by a region of high dielectric. This theoretical approach gives the choice to either include average solution salt effects (PB) or to use the pure solvent (PE). Solving PE for such a system is equivalent to calculating a charge density around the solute surface at the boundary where the dielectric changes.[39] This, among other things, allows the calculation of the free energy of charging of a cavity in a continuum solvent where, at least in the case of water, polarization comes mostly from solvent nuclear motion averaging. While the dielectric boundary is de facto representing the molecular polarization, the dielectric constants and radii employed traditionally are parametrized by fitting to energies (such as solvation or binding free energies) without regard for the molecular polarizabilities themselves. These energies are also dependent on details of the molecular electronic charge distribution, the solvent/solute boundary, and sometimes the nonpolar energy terms, all of which obfuscate the parametrization with respect to the key property of molecular polarizability.

Our approach is to use an intramolecular effective dielectric constant, together with associated atomic radii, to accurately represent the detailed molecular polarizability. For this to be a widely applicable model of polarizability, the generality between related chemical species of a given set of intramolecular effective dielectric constants and associated atomic radii would have to be demonstrated. Such a polarizability model, independent per se of the molecule's charge distribution, could then subsequently be combined with a suitable static charge model to produce a polarizable electrostatic term applicable to force fields.

To evaluate the model, the simplest starting point is gas-phase polarizabilities, using a higher dielectric value inside the molecule and vacuum dielectric outside.[40] This way, the charge density formed at the exterior/interior boundary comes from the polarization of the molecule alone. Comparison of the polarizability tensors from such calculations directly to those from B3LYP/aug-cc-pVTZ calculations allows proof-of-concept of the model. The resulting parameters can be used to rapidly calculate molecular polarizabilities on large molecules.

To calculate the molecular polarizability, we first solve EPIC for a system in which the interior/exterior boundary is described by a van der Waals (vdW) surface, an inner dielectric, and a uniform electric field. The electric field is simply produced from the boundary conditions when solving on a grid (electric clamp). From the obtained solution, it is possible to calculate the charge density from Gauss' law (i.e., from the numerical divergence of the electric field), and the induced dipole moment is simply the sum of the grid charge times its position as shown by eq 1 below.

$$\vec{\mu}^{\text{ind}} = \sum_{i=1}^{\text{grid}} \vec{r}_i \cdot q_i \tag{1}$$

Knowing the applied electric field, it is then possible, as shown in eq 2, to compute the polarizability tensor given that three calculations are done with the electric field applied in orthogonal directions; in eq 2, i and j can be x, y, or z.

$$\alpha_{ij} = \mu_i^{\text{ind}}/E_j \tag{2}$$

3.2. Spherical Dielectric. For the sake of clarifying the internal structure of the model, let us first consider the induced polarization of a single atom in vacuum under the influence of a uniform external electric field: the EPIC model for an atom. Given a sphere of radius R, a unitless inner dielectric ε_{in} and the uniform electric field E, we can exactly calculate the induced dipole moment with eq 3.

$$\vec{\mu}^{\text{ind}} = 4\pi\varepsilon_0\left(\frac{\varepsilon_{\text{in}} - 1}{\varepsilon_{\text{in}} + 2}\right)R^3 \cdot \vec{E} \tag{3}$$

Here, the atomic polarizability is given by the electric field E prefactor, which is a scalar given the symmetry of the problem. The induced dipole moment originates from the accumulation of a charge density at the boundary of the sphere opposing the uniform electric field.[39] From eq 3, we see that the polarizability has a cubic dependency on the sphere radius and that the inner dielectric can reduce the polarizability to zero (ε_{in}=1), while the upper limit of its contribution is a factor of 1 ($\varepsilon_{\text{in}} \gg 1$). The contribution of ε_{in} to the atomic polarizability asymptotically reaches a plateau as shown in Figure 1. Thus, at high values of ε_{in}, the atomic radius becomes the dominant dependency in the

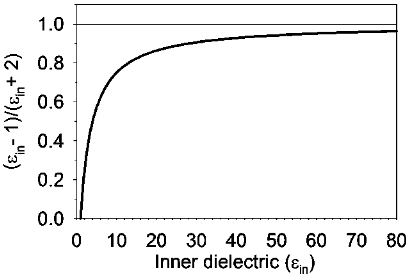


Figure 1. Dielectric contribution to the sphere dielectric continuum polarizability goes asymptotically to one and most of the contributions are below $\varepsilon_{in} = 10$.

electric field prefactor; we find similar characteristics for nonspherical shapes.

It is interesting to make a parallel between eq 3 and the PID model, where the polarizable point would be located exactly at the nucleus. In this particular case, it is possible to equate the polarizability from PE, induced by the radius and the dielectric, to any point polarizability.[11] However, when the electric field is not uniform, the PID induced atomic dipole originating from the evaluation of the electric field at a single point may not be representative, leading to inaccuracies.[41] This is in contrast with the EPIC model that builds the response based on the electric field lines passing locally through each part of the atom's surface, allowing a response more complex than that of a point dipole. In molecules, the atomic polarizabilities of the PID model do not find their counterparts in the EPIC model since it is difficult to assign nonoverlapping dielectric spheres to atoms and obtain the correct molecular behavior. The Cl_2 molecule studied in this work is an example.

4. Methods

4.1. Calculations. Prior to the DFT calculation, SMILES[42−44] strings of the desired structures were transformed into hydrogen-capped three-dimensional structures with the program OMEGA.[45] The *n*-octane conformer set was also obtained from OMEGA. The resulting geometries were optimized with the Gaussian '03[46] program using B3LYP[47−49] with a 6-31++G(d,p) basis set[50,51] without symmetry. The atomic radii and molecular inner dielectrics were fit based on molecular polarizability tensors calculated at the B3LYP level of theory[52] with the Gaussian '03 program. The extended Dunning's aug-cc-pVTZ basis set,[53,54] known to lead to accurate gas phase polarizabilities, was used.[55] An extended basis set is required to obtain accurate gas phase polarizabilities that would otherwise be underestimated.

The solutions to the PE were obtained with the finite difference PB solver Zap[56] from OpenEye Inc. modified to allow voltage clamping of box boundaries to create a uniform electric field. The electric field is applied perpendicularly to two facing box sides (along the *z* axis). The difference between the fixed potential values on the boundaries is set to meet: $\Delta\varphi = E_z \times \Delta Z$, where $\Delta\varphi$ is the difference in potential, E_z is the magnitude of the uniform electric field, and ΔZ is the grid length in the *z* direction. The salt concentration was set to zero, and the dielectric boundary was defined by the vdW surfaces. The grid spacing was set to 0.3 Å, and the extent of the grid was set such that at least 5 Å separated the box wall from any point on the vdW surface. As detailed in the Supporting Information, grid spacing below 0.6 Å did not show significant deterioration of the results. Small charges of ±0.001e were randomly assigned to the atoms to ensure Zap would run, typically converging to 0.000001 kT.

In tables where optimized parameters are reported, a sensitivity value associated with each fitted parameter is also reported. The sensitivity of a parameter corresponds to its smallest variation, producing an additional 1% error in the fitness function considering only molecules using this parameter. The sensitivity is calculated with a three-point parabolic fit around the optimal parameter value, and the change required obtaining the 1% extra error is extrapolated. Therefore, the reported sensitivity indicates the level of precision for a given parameter and whether or not some parameters could be eventually merged.

4.2. Fitting Procedure. Equation 4 shows the fitness function *F* utilized in the fitting of the atomic radii and the inner dielectrics.

$$F(\{R\},\{\varepsilon\}) = \frac{1}{3N}\sum_{i=1}^{N}\sum_{j=xx,yy,zz}\frac{|\alpha_{ij}^{\mathrm{QM}}-\alpha_{ij}|}{\alpha_{ij}^{\mathrm{QM}}} + \frac{1}{N_\theta}\sum_{i=1}^{N_\theta}\frac{1-|\vec{v}_{ij}^{\mathrm{QM}}\cdot\vec{v}_{ij}|}{1-\cos 45°} \qquad (4)$$

In eq 4, *N* corresponds to the number of molecules used in the fit, α_{ij} to the polarizability component *j* of the molecule *i*, and v_{ij} to the eigenvector of the polarizability component *j* of molecule *i*. N_θ is the number of nondegenerate eigenvectors found in all the molecules. This fitness function is minimal when the three calculated polarizability components are identical to the QM values and when the corresponding component directions are aligned with the QM eigenvectors of the polarizability tensor.

As shown in the Cl_2 example of Figure 2, the hypersurface of eq 4 has a number of local minima; it is important that our fitting procedure allows these to be examined. Because the calculations were fast, we decided to proceed in two steps: First, a systematic search was carried out varying each fitted parameter over a range and testing all combinations. The 30 best sets of parameters were then relaxed using a Powell minimization algorithm, and the set of optimized parameters leading to the smallest error was kept.

4.3. Definitions. The polarizability tensor is a symmetric 3 × 3 matrix derived from six unique values. It can be used to calculate the induced dipole moment μ_i (*i* takes the value *x*, *y*, and *z*) given a field vector *E*:

$$\mu_i^{\mathrm{ind}} = \alpha_{ix}E_x + \alpha_{iy}E_y + \alpha_{iz}E_z \qquad (5)$$

In this work, we use the eigenvalues and eigenvectors of the polarizability tensor. The eigenvalues are rotationally invariant, and their corresponding eigenvectors indicate the direction of the principal polarizability components. The three molecular eigenvalues are named α_{xx}, α_{yy}, and α_{zz}, and by convention $\alpha_{xx} \le \alpha_{yy} \le \alpha_{zz}$. The average polarizability (or isotropic polarizability) is calculated with eq 6 below. We also define the polarizability anisotropy

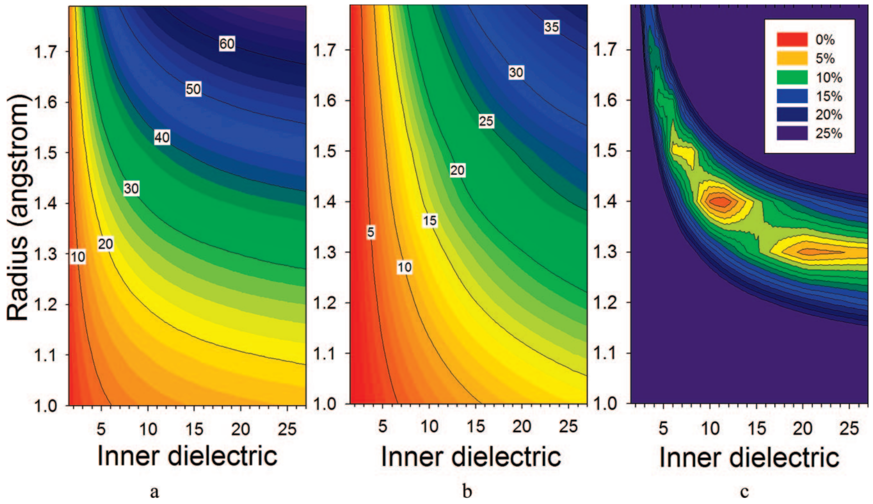


Figure 2. EPIC model behavior is explored for Cl_2. The average polarizability (a) and the anisotropy (b) isolines (in au) are plotted as a function of the Cl atomic radius, used to define the vdW surface, and the value of the inner dielectric. The target Cl_2 B3LYP values are 31.43 (average) and 18.24 (anisotropy) (cf. Table 1). The polarizability tensor error function $(2|\alpha_{\perp}^{QM} - \alpha_{\perp}^{EPIC}| + |\alpha_{\parallel}^{QM} - \alpha_{\parallel}^{EPIC}|)/3\alpha_{avg}^{QM}$ isolines in (c) identify the regions where the EPIC model matches the B3LYP polarizability tensor. The external dielectric is set to one, and the internuclear distance of Cl_2 is fixed at 2.05 Å. These figures show that a high dielectric value is required to match the QM anisotropy and that a number of minima can be found on the error hypersurface.

Table 1. Compared Polarizabilities (au) of Diatomic Molecules when the Radii and ε_{in} Are Fit to B3LYP/aug-cc-pVTZ Polarizabilities[a]

		$\alpha_{\perp}$	$\alpha_{\parallel}$	α_{avg}	$\Delta\alpha$	δ_{avg}[b] (%)	δ_{aniso}[b] (%)
H_2	EPIC (0.88, 7.8)[c]	4.92	6.83	5.55	1.91	0.1	0.3
	(0.83)[d]	4.47	6.60	5.18	2.12	6.7	4.1
	B3LYP	4.92	6.81	5.55	1.89		
	exp[e]	4.86	6.28	5.33	1.42		
N_2	EPIC (1.02, 19.5)[c]	10.49	15.89	12.29	5.40	1.8	3.7
	(1.03)[d]	10.35	15.58	12.09	5.23	0.2	2.3
	B3LYP	10.42	15.38	12.07	4.96		
	exp[e]	9.8	16.1	11.90	6.3		
F_2	EPIC (0.86, 20.5)[c]	6.26	12.64	8.39	6.37	0.5	1.5
	(0.84)[d]	6.06	11.20	7.77	5.14	6.9	16.3
	B3LYP	6.18	12.68	8.35	6.50		
Cl_2	EPIC (1.34, 19.3)[c]	25.64	43.90	31.73	18.26	0.9	0.1
	(1.34)[d]	25.38	43.03	31.26	17.65	0.7	1.9
	B3LYP	25.35	43.59	31.43	18.24		
	exp[e]	24.5	44.6	31.15	20.1		
Br_2	EPIC (1.53, 17.5)[c]	36.84	62.42	45.37	25.57	1.0	2.2
	(1.52)[d]	36.19	62.73	45.04	26.54	1.7	0.1
	B3LYP	36.96	63.53	45.82	26.57		

[a] Two fitting methods are involved: 1 radius and 1 dielectric per element, 1 radius per element, and a single dielectric for all five. [b] Error relative to B3LYP values using eqs 9 and 10 with $N = 1$. [c] The number in the parentheses are the optimal (radius Å, dielectric) individually fit for each molecule. [d] The optimal radius (in Å) fit for each individual diatomic is reported in parentheses given a globally fit dielectric of 18.0. [e] Experimental values are from ref 19.

in eq 7. This particular definition of anisotropy is an invariant in the Kerr effect and has been often used in the literature.[57]

$$\alpha_{avg} = \frac{\alpha_{xx} + \alpha_{yy} + \alpha_{zz}}{3} \tag{6}$$

$$\Delta\alpha = \sqrt{\frac{(\alpha_{xx} - \alpha_{yy})^2 + (\alpha_{xx} - \alpha_{zz})^2 + (\alpha_{yy} - \alpha_{zz})^2}{2}} \tag{7}$$

Equation 7 can be rewritten in terms of only two independent differences in the polarizabilities as shown in eq 8,

$$\Delta\alpha = \sqrt{a^2 + b^2 + ab} \tag{8}$$

where $a = \alpha_{zz} - \alpha_{yy}$ and $b = \alpha_{yy} - \alpha_{xx}$. In the case of degenerate molecules as in diatomics, eq 8 reduces to the unsigned difference between two different polarizability eigenvectors.

We now define errors as used in the rest of this article. Equation 9 gives the average unsigned error of the approximated anisotropy ($\Delta\alpha$) where N corresponds to the number of molecules, $\alpha_{i,avg}$ to the average polarizability (eq 6) of molecule i, and QM corresponds to the DFT values.

$$\delta_{aniso} = \frac{1}{N}\sum_{i=1}^{N} \frac{|\Delta\alpha_i^{QM} - \Delta\alpha_i|}{\alpha_{i,avg}^{QM}} \tag{9}$$

Similarly, the average unsigned error of the average polarizability is defined by

$$\delta_{avg} = \frac{1}{N}\sum_{i=1}^{N} \frac{|\alpha_{i,avg}^{QM} - \alpha_{i,avg}|}{\alpha_{i,avg}^{QM}} \tag{10}$$

Finally, we define an average angle error between the eigenvectors ν from QM and our parametrized model as

$$\theta = \frac{1}{N_\theta}\sum_{i=1}^{N_\theta} |\cos^{-1}(\vec{v}_i \cdot \vec{v}_i^{QM})| \tag{11}$$

We prefer the use of the error in the average polarizabiliy, the anisotropy, and the deviation angle over the error in the polarizability components or the tensor elements. This allows us to analyze the physical origin of the errors and in particular how much comes from anisotropy, normally a more stringent property to fit.

4.4. Molecule Data Sets. Our data set is made to challenge the EPIC model with anisotropic cases known to be difficult. It is formed from three chemical classes:

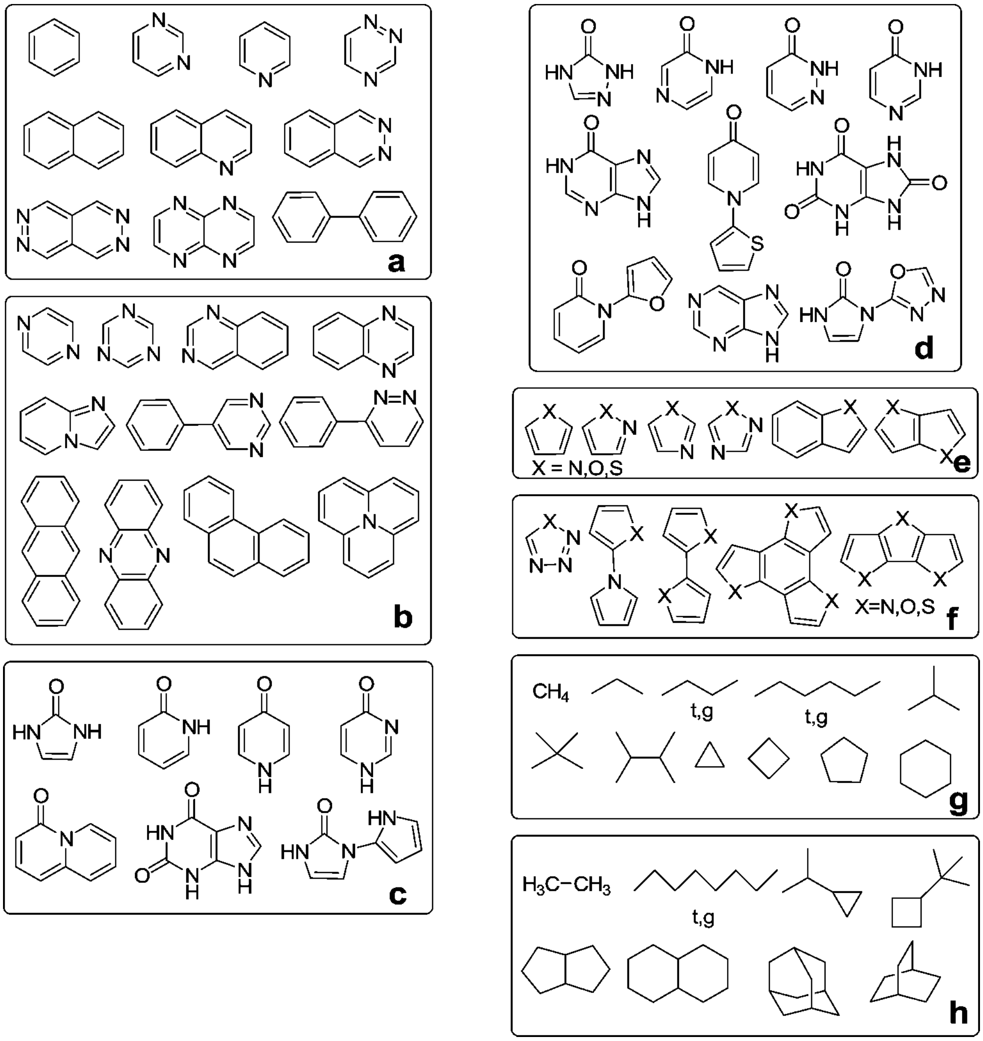


Figure 3. Molecules used are divided in 12 data sets and 6 chemical classes: the heteroaromatics training set “aromatics-t” (a), the heteroaromatics validation set “aromatics-v” (b), the pyridones training set “pyridones-t” (c), the pyridones validation set “pyridones-v” (d), the furans training set “furans-t” (X = O), the pyrroles training set “pyrroles-t” (X = N), the thiophenes training set “thiophenes-t” (X = S) (e), the furans validation set “furans-v” (X = O), the pyrroles validation set “pyrroles-v” (X = N), the thiophenes validation set “thiophenes-v” (X = S) (f), the alkanes training set “alkanes-t” (g), and the alkanes validation set “alkanes-v” (h). The X atoms in a molecule are all O, all S, or all NH. In the case of *n*-butane, *n*-hexane, and *n*-octane, two conformers are considered: all trans (t) and gauche (g).

diatomics, heteroaromatics, and the alkanes. While not comprehensive, these data sets were deemed sufficient for proof of concept. Except for the diatomics, all the molecules examined are subdivided into 12 data sets and 6 chemical classes as in Figure 3. For each class there is a training set (“-t” suffix), used in the parametrization, and a validation set (“-v” suffix) to verify the transferability of the obtained parameters.

Trying to cover a broad range of unsubstituted heteroaromatic molecules, we selected five classes of aromatics: heteroaromatics, pyridones, pyrroles, furans, and thiophenes. The aromatics are limited to C, H, and divalent N atoms. The pyridones contain aromatic amides; while these also exist under their hydroxypyridine tautomers, in water the equilibrium is strongly driven toward the pyridone form, which we exclusively study. The pyrroles, furans, and thiophenes classes are made from the same scaffolds except differing by one atomic element for each class. In the training sets, balancing the number of molecules is important to avoid overfitting. Each nondegenerate molecular polarizability tensor contributes six datapoints (i.e., from six independent tensor elements). Degenerate molecules contribute either four or one independent data points, depending on the degree of symmetry. The pyridones-v, the pyrroles-v, the thiophenes-v, and the furans-v sets all contain multiple functional groups.

The alkanes-t set contains both small and large isotropic molecules (methane and neopentane). It also contains anisotropic molecules like *trans*-hexane. We included two conformers of butane and hexane because their isotropic polarizability is similar but their anisotropy differs. Cyclic species are also included as a result of their special nature. The alkanes-v set contains fused cyclic alkanes and an octane in two different conformations of which the trans form is highly anisotropic. We also mixed cyclic alkanes with chain alkanes in the validation set; all this with the desire of having a validation set significantly different from the training set to really assess the transferability of the fitted parameters.

For this reason, none of the molecules from the validation sets are used in the parametrization.

5. Results

5.1. Diatomics: The Cl_2 Polarizability Hypersurface. The Cl_2 homonuclear diatomic is the simplest molecule that unveils the dependency of the polarizabilities on the radius and the inner dielectric. In Figure 2, parameter hypersurfaces are illustrated for Cl_2 made of two spheres of radius R separated by 2.05 Å (DFT equilibrium distance) within which the inner dielectric is higher than one and the outer dielectric set to the vacuum value of one. When the two spheres overlap ($R > 1$ Å), the molecular volume is described by a vdW surface. Figure 2a shows the contour plot of the average polarizability of the molecule as a function of the Cl radius and inner dielectric. As with the sphere polarizability, the radius has a strong impact on the average polarizability, and the influence of the inner dielectric is significantly reduced beyond a value of 10. The anisotropy, however, is more affected by the dielectric constant and varies less rapidly and over a larger range of radius and dielectric than the average polarizability. The Cl_2 example illustrates the need for high dielectric compared to experimental values, and this is especially true when a molecule is highly anisotropic. Figure 2b shows that for low values of the inner dielectric, the dependence of the anisotropy on the radius diminishes.

Importantly, it is clear that the EPIC model does not have the polarizability catastrophe problem associated with the PID family of polarizable models. When two polarized spheres start to overlap, the interaction between the induced dipoles does not diverge. One reason for this is that the induced polarization is spread over space, rather than being concentrated at a point. Also, when two atoms approach each other, their volumes and, hence, the total polarizability are decreased. Hence, the atomic radii in the EPIC model play a role somewhat similar to the Thole shielding factor used in PID and DO models.

The Cl_2 bond-parallel and -perpendicular polarizabilities obtained by DFT are 25.4 and 43.6 au, respectively, leading to an average polarizability of 31.4 au and an anisotropy of 18.2 au. Pairs of radius and dielectric that can reproduce the DFT values and can be visually identified by plotting the isolines of the fitness function as shown in Figure 2c.

$$F(R, \varepsilon) = \frac{2|\alpha_{\perp}(R, \varepsilon) - \alpha_{\perp}^{QM}| + |\alpha_{||}(R, \varepsilon) - \alpha_{||}^{QM}|}{3\alpha_{avg}^{QM}}$$

Four local minima are identified (three are obvious from the figure) from which two, located at ($R = 1.4$, $\varepsilon = 11.5$) and ($R = 1.3$, $\varepsilon = 20.0$) produce an overall error less than 5%. The existence of the multiple minima is due to the multi-objective nature of the fitness function: the error surface has minima where the isolines of ~30 au in Figure 2a and the isoline of ~20 au in Figure 2b are close to each other, simultaneously matching the DFT values. Higher minima are found when only one of the anisotropy or the average polarizability match the DFT values. For instance, at ($R = 1.5$, $\varepsilon = 7.0$) the average value is matched but not the anisotropy. Similar hypersurfaces have been found with PE in a different context.[37,58]

Finally, it is interesting to note, as alluded to in the previous section, that for Cl_2 it is not possible to assign a small sphere (<1 Å) to each atom, no matter how large the dielectric, and reproduce the correct polarizability. This clarifies the difference between the EPIC and the PID models. Although they both serve the same purpose, the two models do not present identical physical pictures. For instance, shielding must be introduced explicitly in PID, whereas it is intrinsic to the physics of the EPIC model.

5.2. Diatomics: Polarizability. Homonuclear diatomic molecules constitute a difficult test for a polarizable model. For example, the FQ model does not allow for bond-perpendicular polarizability, which is typically half of the bond-parallel polarizability. van Duijnen et al.[14] have reparameterized the PID-Thole model, and they obtained 22% error on the average polarizabilities of H_2, N_2, and Cl_2. Their error in the anisotropy is significantly larger. More recently, a special parametrization for homohalides with the PID-Thole model gave errors of 9% and 82% on the average polarizability and anisotropy of F_2, respectively.[13] In the case of Cl_2, the errors on the average polarizability and anisotropy are 2% and 20%; finally, for Br_2 the same authors found 0.8% and 13%. However, Birge[20] assigned anisotropic atomic polarizabilities and obtained the experimental values for H_2 and N_2. These large errors of the models without atomic anisotropy corrections have been attributed to the difficulty of increasing the atomic induced dipole interaction. Fitting our model to match B3LYP/aug-cc-pVTZ molecular polarizabilities led to significantly smaller errors as shown in Table 1. In the best case, we fit a different inner dielectric and radius for each element. This is a good example of overfitting since two parameters are used to reproduce two polarizabilities. However, it is a way to verify that the dielectric model is flexible enough to deal with the diatomics without using atomic anisotropy parameters. Table 1 shows the results for five diatomic molecules, and the reported errors for the average polarizability and anisotropy are 0.1% and 0.3% for H_2, 1.8% and 3.7% for N_2, 0.5% and 1.5% for F_2, 0.9% and 0.1% for Cl_2, and 1.0% and 2.2% for Br_2. These results clearly show enough flexibility to account for both average polarizability and anisotropy. The second fitting scenario involved a single dielectric for all five molecules and five atomic radii, fitting 6 parameters to 10 data points. The optimal parameters give results still in relatively good agreement with DFT with a maximum of 16% error made in the case of F_2 anisotropy. For both optimal parameter sets, the radii and dielectrics are reported in Table 1 in parentheses.

These encouraging results on diatomics show that the EPIC model can correctly account for polarizability on a minimal group of two atoms. Therefore, we expect that the local polarizability may be well represented in larger molecules.

5.3. Organic Data Sets: Typical PB Parameters. As an initial check on how well typical radii and inner dielectric used in PB applications could reproduce the molecular polarizabilities, we first examined the set of parameters obtained by Tan and Luo[12] that lead to reasonable dipole moments in different continuum external dielectrics. In their

Table 2. Unsigned Average Errors for All Molecules in Figure *3*, Relative to B3LYP/aug-cc-pVTZ, of Average Polarizability and Anisotropy Obtained with Various Parameters Typically Used in Pb Applications

radii	ε_{in}	δ_{avg} (%)	STDEV (%)	δ_{aniso} (%)	STDEV (%)
Tan and Luo[a]	4	52	20	18	10
CHARM22[b]	2	40	13	47	23
	4	26	26	28	13
	8	84	40	17	26
	16	129	50	54	44
Bondi[c]	2	51	6	47	23
	4	9	6	26	15
	8	51	15	14	16
	16	91	17	52	29
EPIC/P2E[d]	4.98, 14.55	2	2	5	4
EPIC/P1E[d]	11.7	2	2	6	6

[a] Reference 12. [b] Reference 59. [c] Bondi radii from reference 60. The Hydrogen radius is set to 1.1 Å following Rowland and Taylor's recommendations.[71] [d] EPIC used with parameters fit in this work as reported in Table 3.

work, they not only fit the inner dielectric but also the atomic charges. They use the PCM radii and obtained a best inner dielectric of 4. This combination of parameters produces an error of 52% in the average polarizability (eq 10) compared to B3LYP (all molecules from Figure 3) and an error of 18% (eq 9) in the anisotropy as outlined in Table 2. In both cases, the standard deviations (STDEV) of the errors are large. The other two sets of radii examined are those from CHARM22[59] and Bondi.[60] We applied four representative inner dielectrics: 2, 4, 8, and 16, spanning the range of dielectrics often reported to be optimal. Table 2 shows very high errors for all the combinations, the best being Bondi radii with an inner dielectric of 4 which led to an average polarizability error of 9% with a STDEV of 6% and an anisotropy error of 26% with a STDEV of 15%. These particular parameters have a bimodal error distribution producing smaller errors for alkanes than for aromatics, which is consistent with other findings (vide infra). Clearly, the parameters from previous studies are not appropriate for the calculation of vacuum molecular polarizabilities, and they do not accurately account for the electronic polarization. When attempting to only optimize the inner dielectric, while keeping the atomic radii to their Bondi values, it was not possible to obtain small errors on the anisotropy.

In the next sections, we present details about new parametrizations that are in much better agreement with DFT values. As outlined in Table 2, we reduced the error produced by the best Bondi combination by a factor of 4 for both the average polarizability and the anisotropy. The STDEV is also greatly reduced allowing for more confidence and robustness in the polarizability predictions.

5.4. Alkanes and Aromatics. Figure 4a,b summarizes the results obtained with the best parameter set, fitted with two inner dielectrics (P2E), for the 12 sets formed by the 6 classes: alkanes, aromatics, pyridones, pyrroles, furans, and thiophenes. The optimal parameters with the atom-typing scheme used to generate the molecular polarizabilities are given in Table 3, along with Bondi radii.[60] In Figure 4, the comparisons are between the DFT polarizabilities and the EPIC model. The errors are reported with histograms and error bars corresponding to the average unsigned errors (eqs 9−11) and the corresponding STDEV indicating the range of variation of the errors.

In Figure 4a, the error on the average polarizabilities is less than 3% for all classes of the training sets and less than 1% for the thiophenes-t set, and the combined average error is less than 2%. The corresponding error on the average polarizabilities for the validation sets in Figure 4b is slightly higher with a maximum of 3.2% for the pyrrole-v set; the combined error is 2.4%.

While this low level of error obtained in the average polarizability has also been observed with other polarizable methods, the anisotropy of the polarizability is less tractable. To capture anisotropy, previous models normally require the use of directional atomic polarizabilities[15,20,21] especially for aromatics. In our training sets, as shown in Figure 4a, we obtain a combined error for the anisotropy of 4%. The worst set, pyridones-t, has an average error of only 7.1%. Although this class is found in biologically active molecules, we could not find published results from other empirical polarizable models for molecular polarizability tensors. We believe that this class might be particularly difficult due to variable aromaticity and accounting for a range of chemical functionalities with the same parameters (imidazolones, 2-pyridones, 4-pyridones, etc.).

The anisotropy average error on the validation set in Figure 4b ranges from 2.5% for the alkanes-v up to 7.4% for the aromatics-v. It is not surprising that the error is larger for the validation sets than for the training sets. Overall, however, when comparing the anisotropy error made on the combined sets, it is not significantly higher: 5.3% for the validation sets versus 4% for the training sets. On the other hand, the STDEV is significantly higher in the validation set.

The aromatics class shows the highest anisotropy shift from the training set to the validation set. Phenazine and phenanthrene are responsible for two out of three large discrepancies between B3LYP and EPIC. It is interesting to note that when comparing B3LYP average polarizability and anisotropy to experiment, the errors are 11% and 30% for phenazine and 17% and 20% for anthracene. The same errors, when comparing our model and experiment, are 5% and 15% for phenazine and 1.7% and 1.4% for anthracene. The EPIC model is thus more accurate for these molecules, which can be partly explained by the known size-consistency defect of DFT for oligocenes (benzene, naphthalene, anthracene, tetracene, etc.) that are usually too anisotropic.[55] In general, DFT methods have problems reproducing the polarizability of long delocalized molecules, and this has been attributed to deficiency of the currently used functionals to account for a self-interaction correction.[61] It is therefore possible that our model, fit on smaller molecules, tends to produce better behavior on these large delocalized molecules. Another implication is that large molecules should not be used for the training of a polarizable model to fit DFT polarizabilities. Figure 5a shows that in fact the correlation between the polarizability components of the entire set of molecules of Figure 3 is excellent up to 150 au. Part of the discrepancy

Table 3. Optimized Radii (Å) and Inner Dielectrics with Sensitivity[a] Accounting for All Molecule Sets (Figure 3): Parameter Sets P2E and P1E

atom type description	optimal value (P2E)	sensitivity	optimal value (P1E)	sensitivity	Bondi radii[b]
		alkanes			
C alkyl	1.39	0.04	1.13	0.03	1.70
H bond on an alkyl C	0.99	0.02	0.78	0.05	1.20
Dielectric alkanes	4.98	0.27	11.70	1.18	
		aromatics			
C aromatic	1.32	0.05	1.30	0.04	1.70
H bonded to aromatic C or N	0.64	0.09	0.78	0.05	1.20
N aromatic	1.06	0.16	1.10	0.14	1.55
O furan-like aromatic	0.74	0.23	0.75	0.27	1.52
O in pyridone carbonyl	0.95	0.25	1.03	0.16	1.52
S thiophene-like	1.50	0.06	1.58	0.05	1.80
dielectric aromatics	14.56	1.50	11.70	1.18	

[a] Smallest parameter variation required to produce a 1% additional error in the fitting function (see Method section for details). [b] Reference 60.

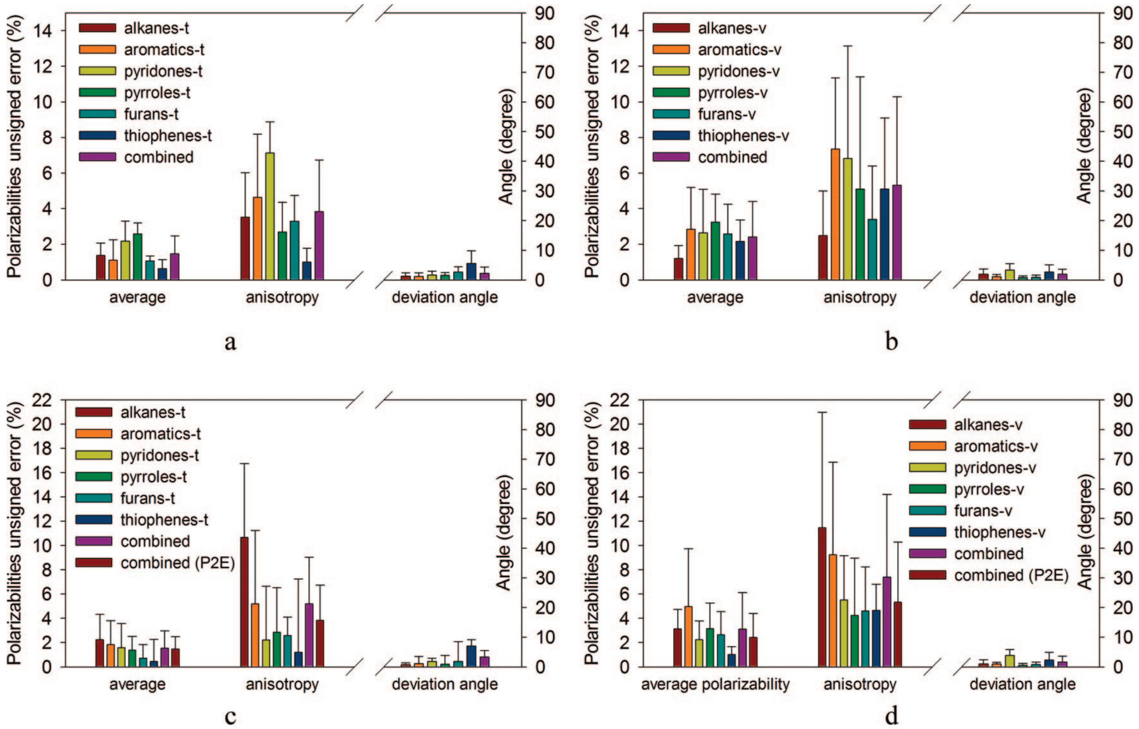


Figure 4. Comparison between B3LYP/aug-cc-pVTZ polarizabilities and EPIC models P2E and P1E for all molecules from Figure 3. The averaged relative error on average polarizability (eq 10), anisotropy (eq 9), and the deviation angle of the eigenvectors (eq 11) are shown together with the corresponding STDEV reported as error bars. The results for the 2-dielectric fit (P2E) training sets (a) and validation sets (b) show small errors in the average polarizability and relatively small errors in the anisotropy. The results for the 1-dielectric fit (P1E) training sets (c) and the validation sets (d) show larger errors in the alkanes anisotropy and generally larger errors than the P2E parameters (shown under combined P2E). Combined errors of the training and validation sets are similar.

might be attributable to a different behavior of DFT methods in that range of polarizabilities. In this respect, optimized effective potential (OEP) and time-dependent DFT methods have shown significant improvement,[62−64] but these are still considerably more resources-intensive. The third worst anisotropy discrepancy between B3LYP and EPIC of this aromatics-v set comes from the cycl[3.3.3]azine molecule which has already shown differences with regular polyacenes in terms of excited states.[65] The transferability for that particular molecule is good, all things considered, with an average polarizability error of 8.6% and anisotropy error of 12.8%.

The pyridones-v set is the most challenging with the highly functionalized purine derivates (purine, hypoxanthine, and uric acid) and the substituted pyridones with five-membered heteroaromatic rings. For example, the geometry optimized 1-(2-thienyl)-pyridin-4-one shows an angle of 58° between the two aromatic rings as opposed to the 1-(oxadiazol)-imidazolone that has the two connected rings coplanar and a fully delocalized electron π system. This data set is similar to the chemical functionalization of drug-like molecules.

The average angles between the eigenvector of the polarizability components of B3LYP and that of the EPIC are less than 5.5° in all sets, although in some molecules

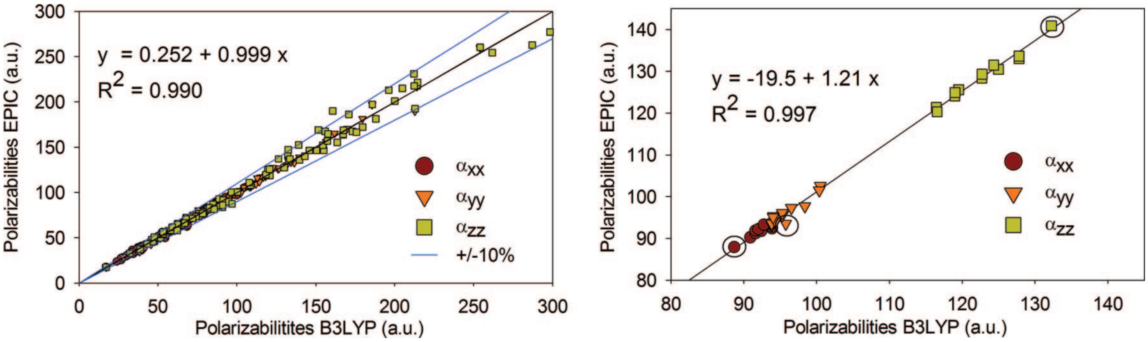


Figure 5. Correlation between B3LYP/aug-cc-pVTZ polarizability components and the EPIC model P2E. In (a), the polarizability components for all sets of Figure 3 are correlated and the ±10% error lines are illustrated. The linear regression shows excellent agreement, especially for polarizabilities smaller than 150 au. In (b), 13 stable conformers of *n*-octane are examined. The all trans conformation polarizabilities are identified with circles. The average polarizability error on the 13 conformers is 1.9%, and the anisotropy error is 5.8%. A linear regression gives an R^2 of 0.997, a slope of 1.21, and an ordinate at the origin of −19.5. This means that the EPIC model P2E overestimates the polarizability of *n*-octane consistently through all conformers.

Table 4. Average Errors and Standard Dveiations (stdev) against Experiment[a] for All Molecules in Figure *3*

method	δ_{avg} (%)	stdev (%)	δ_{aniso} (%)	stdev (%)
Tan and Luo[b]	58.4	19.8	13.6	9.4
Bondi[c]	8.3	6.2	22.4	13.5
EPIC/P2E[d]	3.9	4.1	9.0	9.5
EPIC/P1E[d]	3.8	3.1	7.3	6.4
B3LYP	4.1	4.1	10.5	9.9

[a] Twenty-five experimental average polarizabilities and 18 anisotropy data. Details given in Supporting Information. [b] Reference 12. [c] Bondi radii and ε_{in} = 4. [d] EPIC used with parameters fit in this work reported in Table 3.

the angles can be as large as 23°, that is, for thiazole. For the pyridones-t and pyridones-v sets, the angular diffences remain surprisingly small.

Finally, Table 4 shows that, compared to experimental values, the parametrized EPIC method performs comparably to B3LYP against the subset of 25 molecules for which experimental data is available. Indeed, EPIC produces a δ_{avg} of 3.9% with experiment compared to 4.1% for B3LYP. It also gives a δ_{aniso} of 9.0% with experiment compared to 10.5% in the case of B3LYP. The STDEV of the errors from B3LYP match EPIC values. The discrepancy between B3LYP and EPIC calculated for the molecules of Figure 3 is smaller leading to a δ_{avg} of 1.9% and a δ_{aniso} of 4.6%. The level of error compared to experiment obtained with both B3LYP and EPIC is not necessarily beyond experimental uncertainty.

5.5. Conformational Dependency of Polarizability. Although we avoided comparing the polarizability of flexible molecules to experimental data, it is obvious that a good empirical method should account for the conformational dependency of the polarizability, the anisotropy, and the orientation of the polarizability tensor eigenvectors. In addition to the deliberate choice of a wide range of 3D diversity in our molecular sets, we examined the case of *n*-octane, the most flexible molecule of the sets. Taking 13 diverse B3LYP geometry optimized conformers of *n*-octane, we computed the polarizability, anisotropy, and the eigenvectors using the P2E parameters. The EPIC method gives average polarizability error and anisotropy error of 1.9% and 5.8%, respectively. Figure 5b shows a correlation graph between B3LYP polarizability components and our model (α_{xx}, α_{yy}, α_{zz}). The correlation is perfectly linear as shown by a linear regression leading to an R^2 of 0.997 although the slope of the regression is 1.21, consistent with the average errors outlined above. Moreover, in Figure 5a, we clearly see that correlation of the polarizability components for all the molecules of Figure 3 is excellent with a slope of 1 and an R^2 of 0.990. This result leads to the conclusion that our model is at least consistently making the same errors for *n*-octane conformers compared to B3LYP. Finally, the orientations of the polarizability components differ by 0.97° with a maximum value of 3.7°; this is in spite of the broken symmetry in the gauche octane conformers.

6. Discussion

6.1. Transferability. Shanker and Applequist,[15] with a variation of the PID model, studied seven nitrogen heterocyclic molecules that we also included in our sets: pyridine, pyrimidine, pyrazine, 9H-purine, quinoxaline, quinoline, and phenazine. Using 12 parameters including directional atomic polarizabilities, they show an average polarizability (eq 10) and anisotropy errors (eq 9) of 10% and 12%, respectively;[66] the parametrized EPIC (Table 3) produces correspondingly 3% and 5% error with only 4 parameters; we feel that the reduced requirement for fitted parameters is due to a better physical model. Similar comparisons can be made to the work of Miller[21] where it is reported that 6 parameters for benzene, 9 parameters for pyridine, 9 parameters for naphthalene, and 12 parameters for quinoline are needed to obtain both the average polarizability and the anisotropy. With the EPIC method, again the same 4 parameters do for all.

Recently, Williams and Stone[67] have parametrized a polarizable model on *n*-propane, *n*-butane, *n*-pentane, and *n*-hexane in both their trans and gauche conformations. With their simplest Ctg model, they use 10 atomic polarizability parameters to fit the polarizability tensors to B3LYP values. They obtain a very small error on both the average polarizability and the anisotropy of 1.16% and 2.37%, respectively. Making the same comparison with our model, we obtain 1.7% average polarizability error and 3.99% anisotropy error. Although the error is slightly larger with our EPIC model, this is obtained with only three parameters also producing similar levels of errors in our extended set of alkanes. Furthermore, the level of errors reported by Williams et al. and our studies are all within the accuracy of B3LYP method.

The small number of parameters (cf. Table 3) needed to fit all the aromatic compounds of Figure 3 is a good indication of the transferability and the generality of the method for heteroaromatic compounds. For example, the same nitrogen radius could simultaneously fit pyridine, pyridone, pyrrole, and even branched nitrogen. In the case of alkanes, we have examined most characteristic shapes. Moreover, the training and validation sets produce similar errors; thus, the expected performance of our method in the general case can be approximated by the errors on the validation sets.

Overall, we obtain the same level of error as the best PID methods parametrized with anisotropic atomic polarizabilities and about threefold more parameters. Although the number of parameters is not an issue for a small and homogeneous set of molecules, it would become a serious barrier for further development of a model applicable to the immense functional group complexity of drug-like molecules, one of the main goals of this ongoing effort.

6.2. Inner Dielectrics. The choice of fitting two inner dielectrics, one for the alkanes and one for the heteroaromatics, makes the calculation of new mixed molecules such as t-butylbenzene not possible unless we have a way to switch from a high dielectric (benzene) to a lower dielectric (t-butyl) intramolecularly. Overall, the value of multiple dielectrics, based on chemical constituency, seems proven as well as being physically reasonable. This is a potentially useful strategy in the development of a future general polarizability model. However, simultaneously fitting the polarizabilities of all the compounds from Figure 3 with a single dielectric still gives reasonable results. Table 3 reports the values of the optimal parameters used to produce the data of Figure 4c,d. We fit one radius per element except for oxygen, which is split into furan-like and pyridone-like, and for carbon which is split into alkane and aromatic. We first had two hydrogen radii, but there was no significant cost to merge them into one single radius. The results, shown in Figure 4c,d, when compared with those of Figure 4a,b, show a significant increase in the errors on the alkanes-t and alkanes-v sets although the errors on the heteroaromatics classes remain similarly small. It is nevertheless surprising that the level of error remains low when describing the electronic dielectric with a single constant when, in principle, the electronic local polarization should vary intramolecularly as suggested by Oxtoby.[68]

Finally, it is reassuring that the best radii for both reported parametrizations follow the chemical sense of atomic size. The remarkably reduced size of the optimal radii compared to conventional vdW radii (like Bondi) is worth few comments. First, the EPIC radii explain a different physics than conventional vdW radii: the latter relate to the repulsive forces that keep molecules apart whereas the former relate to the electronic response inside the molecule. There is no reason a priori that they would be the same. Furthermore, the high dielectric and the small radii are necessary to modulate the molecular shape so as to correctly fit the polarizability anisotropy. For example, a benzene molecule is flattened when the carbon radii are reduced, and thus the out-of-the-plane polarizable volume is reduced while the in-the-plane length is more or less conserved, increasing the anisotropy. With smaller radii reducing the molecular volume for dielectric response, a higher dielectric value is then needed to conserve the molecular polarizability (cf. eq 3).

6.3. Link to the Optical Dielectric Constants. Intramolecular dielectric constants in the context of PE or PB can adopt many values depending on the system and the phenomena involved[35,37,58,69] and have been attributed values from 1 to 20. The optimal inner dielectric of solutes in continuum solvent free energy and in ligand−protein binding calculations do not agree.[37] Here, we attempt to position our work in this jungle of dielectrics.

We are concerned uniquely with the electronic polarization component. None of the optimal dielectric constants fitted in this work match the experimental optical dielectric constants calculated as the square of the refractive index, which normally have values between 1.2 and 4.0. We partly justify the need for larger dielectrics in section 6.2, but there are other factors that should also be considered. It is important to realize that the link between the molecular polarizability and the macroscopic optical dielectric constant is given by the Lorentz−Lorenz relation shown in eq 12 where N is the number of molecules in the volume V and ε is the macroscopic dielectric when the light frequency is high compare to the dipolar or ionic relaxation time (ε_0 is the vacuum permittivity constant).

$$\alpha_{\mathrm{avg}} = \frac{3\varepsilon_0 V}{N}\left[\frac{\varepsilon - 1}{\varepsilon + 2}\right] \tag{12}$$

In the Lorentz−Lorenz equation a molecule is approximated as a spherical dielectric with an effective molecular volume given by the ratio of the macroscopic space occupied by one molecule. However, from our atomistic perspective the effective volume of a molecule is defined by the electronic density and does not include the empty space between molecules effectively included in eq 12. Hence, in the EPIC model that we parametrize, the average polarizability is the link to the refractive index and not the inner dielectric. The main reason for this is the inconsistency between the atomistic and the macroscopic definitions of the molecular volume. This raises the point that using experimental optical dielectrics assigned to the solute interior in continuum solvent approaches should be further questioned.

Finally, we believe that a more accurate treatment of solute polarizability in the context of continuum solvent could improve the quality of continuum dielectric methods. Obviously the radii and dielectrics obtained in the present work cannot be used in the condensed phase directly; conventional vdW radii should be used as the basis for intermolecular contacts (such as hydrogen bonding) and the solvent boundary. Therefore, to simultaneously include the solute electronic response and the correct solvent response, there is a dielectric region, which still needs to be characterized, in between our small "polarizability" radii and the vdW radii. Although out of scope for the present article, we are in the process of extending the use of our findings in this direction. Once done, one could think of obtaining a polarizable model close to the "polarizable continuum model" (PCM) of Tomasi et al.[70] in which the electronic density would be simply replaced by an "electronic volume" defined with radii and a dielectric constant.

7. Conclusion

In this work, the simple physical picture afforded by a continuum dielectric representation has been used to accurately model molecular dipole polarizability tensors. The molecular inner dielectric in the EPIC model accounts for the electronic polarization. To tackle gas-phase polarizabilities, we capitalized on existing finite difference Poisson−Boltzmann code to calculate the induced dipole moment of a molecule in vacuum in the presence of a uniform electric field. As opposed to the usual use of PE or PB in continuum models, the molecule is a region of higher dielectric and the external dielectric is set to the vacuum value. The calculations are fast and resource-sparing, with equivalently good results up to a grid spacing of 0.5 Å, even though a discrete vdW dielectric boundary is used.

This EPIC model of molecular polarizability possesses some important differences with other approximations such as the point inducible dipole, Drude oscillator, and the fluctuating charge models. It is based on a local differential equation solved on a grid, which brings to the same level of complexity the polarizability and Coulombic electrostatic components. Importantly, EPIC avoids the polarizability catastrophe found in the other PID-based models. Furthermore, it allows, in principle, for a more detailed response to the electric field than the PID or the FQ models based on the fact that the response emerges from the electric field lines across the molecule surface instead of evaluations only at atomic nuclear positions.

This study involved the parametrization of atomic radii, used in the definition of the vdW dielectric boundary, and the molecular inner dielectric. Previous values of these parameters found in the literature are unacceptably poor at approximating molecular polarizability, especially the anisotropy. We attribute this discrepancy to the fact that previous models simultaneously optimize different kinds of interdependent parameters fitting to a complex energy property instead of focusing on solute polarization. Indeed, the previous purpose of using dielectric continuum was in the context of continuum solvent, often completely neglecting the solute response per se.

To test the newly proposed method, we selected difficult chemical classes: the homonuclear diatomics, a wide variety of heteroaromatics, and a diverse set of alkanes. A total of 5 diatomics plus 48 molecules are part of the training sets, subdivided into 6 chemical classes to which we add 45 molecules for validation purposes. In previous models, the polarizabilities of these classes of compounds were correctly calculated only when anisotropic atomic polarizabilities were employed (or auxiliary sites in the case of FQ). Already, with about threefold less parameters than other studies with different models, we have obtained averaged polarizability errors smaller than 5% and averaged anisotropy errors less than 8% considering all sets. The polarizability components calculated with the EPIC/P2E model correlates very well with B3LYP/aug-cc-pVTZ with an R^2 of 0.990 and a slope of 0.999. The orientations of the polarizability eigenvectors are also well reproduced. The flexibility of the model even allowed the calculation of an accurate anisotropy for F_2 without resorting to auxiliary sites or anisotropic parameters. We also found that the EPIC model was able to consistently calculate the molecular polarizabilities on 13 different conformers of *n*-octane. Because of the success of parsimonious parametrization of the EPIC model on difficult chemical classes, we believe that the parametrization can be generalized for all organic chemistry with adequate accuracy. In doing this, we found that intramolecularly varying dielectric constant might be needed to account for the molecular anisotropy.

Overall, this study exemplified that a phenomenon as complex as electronic polarization can be accurately modeled with a simple dielectric continuum model. The principal implications of these findings are in the areas of Poisson−Boltzmann methods and in polarizable force field development. However, the level of accuracy obtained might also have impact beyond our initial consideration, for example, in the field of spectroscopy.

Acknowledgment. Sathesh Bhat from Merck Frosst Canada Ltd. is thanked for helpful comments on the manuscript. Roger Sayle from OpenEye Inc. provided IUPAC names for the most challenging molecules. This work was made possible by the computational resources of the réseau québécois de calcul haute performance (RQCHP). The authors are grateful to OpenEye Inc. for free academic licenses. R.I.I. acknowledges financial support from the Natural Sciences and Engineering Research Council of Canada (NSERC). J.-F.T. is supported by NSERC through a Canada graduate scholarship (CGS D) and by Merck & Co. through the MRL Doctoral Program I. B.R. is supported by NIH Grant GM072558.

Supporting Information Available: DFT, experimental, and EPIC polarizabilities are available for all molecules examined. The optimized coordinates of all molecules are also included. Further discussion on grid spacing is included. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Lane, N. F. *Rev. Mod. Phys.* **1980**, *52*, 29–119.

(2) Kirkwood, J. G. *J. Chem. Phys.* **1937**, *5*, 479–491.

(3) Wagnière, G. H. *Linear and Nonlinear Optical Properties of Molecules*, VCH ed.; Helvetica Chimica Acta Publishers: Weinheim, 1993.

(4) Maroulis, G.; Hohm, U. *Phys. Rev. A* **2007**, *76*, 032504.

(5) Vela, A.; Gazquez, J. L. *J. Am. Chem. Soc.* **1990**, *112*, 1490–1492.

(6) Nagle, J. K. *J. Am. Chem. Soc.* **1990**, *112*, 4741–4747.

(7) Allen, T. W.; Andersen, O. S.; Roux, B. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 117–122.

(8) Wick, C. D.; Kuo, I. F. W.; Mundy, C. J.; Dang, L. X. *J. Chem. Theory Comput.* **2007**, *3*, 2002–2010.

(9) Guo, H.; Gresh, N.; Roques, B. P.; Salahub, D. R. *J. Phys. Chem. B* **2000**, *104*, 9746–9754.

(10) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. *Adv. Drug Delivery Rev.* **1997**, *23*, 3–25.

(11) Sharp, K.; Jean-Charles, A.; Honig, B. *J. Phys. Chem.* **1992**, *96*, 3822–3828.

(12) Tan, Y. H.; Luo, R. *J. Chem. Phys.* **2007**, *126*, 094103.

(13) Elking, D.; Darden, T.; Woods, R. J. *J. Comput. Chem.* **2007**, *28*, 1261–1274.

(14) van Duijnen, P. T.; Swart, M. *J. Phys. Chem. A* **1998**, *102*, 2399–2407.

(15) Shanker, B.; Applequist, J. *J. Phys. Chem.* **1996**, *100*, 3879–3881.

(16) Silberstein, L. *Philos. Mag.* **1917**, *33*, 521–533.

(17) Bode, K. A.; Applequist, J. *J. Phys. Chem.* **1996**, *100*, 17820–17824.

(18) Applequist, J. *J. Phys. Chem.* **1993**, *97*, 6016–6023.

(19) Applequist, J.; Carl, J. R.; Fung, K. K. *J. Am. Chem. Soc.* **1972**, *94*, 2952–2960.

(20) Birge, R. R. *J. Chem. Phys.* **1980**, *72*, 5312–5319.

(21) Miller, K. J. *J. Am. Chem. Soc.* **1990**, *112*, 8543–8551.

(22) Thole, B. T. *Chem. Phys.* **1981**, *59*, 341–350.

(23) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227–249.

(24) Cieplak, P.; Kollman, P. A.; Lybrand, T. *J. Chem. Phys.* **1990**, *92*, 6755–6760.

(25) Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A. *J. Phys. Chem. A* **2004**, *108*, 621–627.

(26) Noskov, S. Y.; Lamoureux, G.; Roux, B. *J. Phys. Chem. B* **2005**, *109*, 6705–6713.

(27) Lamoureux, G.; Roux, B. *J. Chem. Phys.* **2003**, *119*, 3025–3039.

(28) Gasteiger, J.; Marsili, M. *Tetrahedron Lett.* **1978**, 3181–3184.

(29) Rick, S. W.; Stuart, S. J.; Berne, B. J. *J. Chem. Phys.* **1994**, *101*, 6141–6156.

(30) Rappe, A. K.; Goddard, W. A. *J. Phys. Chem.* **1991**, *95*, 3358–3363.

(31) Chelli, R.; Procacci, P.; Righini, R.; Califano, S. *J. Chem. Phys.* **1999**, *111*, 8569–8575.

(32) Harder, E.; Anisimov, V. M.; Whitfield, T.; MacKerell, A. D.; Roux, B. *J. Phys. Chem. B* **2007**, *112*, 3509–3521.

(33) Honig, B.; Nicholls, A. *Science* **1995**, *268*, 1144–1149.

(34) Roux, B.; MacKinnon, R. *Science* **1999**, *285*, 100–102.

(35) Antosiewicz, J.; McCammon, J. A.; Gilson, M. K. *J. Mol. Biol.* **1994**, *238*, 415–436.

(36) Simonson, T.; Archontis, G.; Karplus, M. *J. Phys. Chem. B* **1999**, *103*, 6142–6156.

(37) Naim, M.; Bhat, S.; Rankin, K. N.; Dennis, S.; Chowdhury, S. F.; Siddiqi, I.; Drabik, P.; Sulea, T.; Bayly, C. I.; Jakalian, A.; Purisima, E. O. *J. Chem. Inf. Model.* **2007**, *47*, 122–133.

(38) Fogolari, F.; Brigo, A.; Molinari, H. *J. Mol. Recognit.* **2002**, *15*, 377–392.

(39) David, J. G. *Introduction to Electrodynamics*, 3rd ed.; Prentice-Hall, Inc.: Upper Saddle River, NJ, 1999.

(40) Nicholls, A. Presented at The 233rd ACS National Meeting, Chicago, IL, March 25−29, 2007.

(41) Schropp, B.; Tavan, P. *J. Phys. Chem. B* **2008**, *112*, 6233–6240.

(42) Weininger, D. *J. Chem. Inf. Model.* **1990**, *30*, 237–243.

(43) Weininger, D.; Weininger, A.; Weininger, J. L. *J. Chem. Inf. Model.* **1989**, *29*, 97–101.

(44) Weininger, D. *J. Chem. Inf. Model.* **1988**, *28*, 31–36.

(45) *OMEGA*, version 2.2.1; OpenEye Scientific Software, Inc.: Santa Fe, NM, 2007.

(46) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N. ; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; 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 C02; Gaussian Inc.: Wallingford, CT, 2004.

(47) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

(48) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372–1377.

(49) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623–11627.

(50) Frisch, M. J.; Pople, J. A.; Binkley, J. S. *J. Chem. Phys.* **1984**, *80*, 3265–3269.

(51) Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer, P. V. *J. Comput. Chem.* **1983**, *4*, 294–301.

(52) Rice, J. E.; Handy, N. C. *J. Chem. Phys.* **1991**, *94*, 4959–4971.

(53) Woon, D. E.; Dunning, J. *J. Chem. Phys.* **1993**, *98*, 1358–1371.

(54) Kendall, R. A.; Dunning, J.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796–6806.

(55) Hammond, J. R.; Kowalski, K.; deJong, W. A. *J. Chem. Phys.* **2007**, *127*, 144105.

(56) Grant, J. A.; Pickup, B. T.; Nicholls, A. *J. Comput. Chem.* **2001**, *22*, 608–640.

(57) Kassimi, N. E. B.; Lin, Z. J. *J. Phys. Chem. A* **1998**, *102*, 9906–9911.

(58) Rankin, K. N.; Sulea, T.; Purisima, E. O. *J. Comput. Chem.* **2003**, *24*, 954–962.

(59) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.;

Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(60) Bondi, A. *J. Phys. Chem.* **1964**, *68*, 441–451.

(61) Sekino, H.; Maeda, Y.; Kamiya, M.; Hirao, K. *J. Chem. Phys.* **2007**, *126*, 014107.

(62) van Faassen, M.; de Boeij, P. L. *J. Chem. Phys.* **2004**, *120*, 8353–8363.

(63) van Faassen, M.; Jensen, L.; Berger, J. A.; de Boeij, P. L. *Chem. Phys. Lett.* **2004**, *395*, 274–278.

(64) van Faassen, M. *Int. J. Mod. Phys. B* **2006**, *20*, 3419–3463.

(65) Leupin, W.; Berens, S. J.; Magde, D.; Wirz, J. *J. Phys. Chem.* **1984**, *88*, 1376–1379.

(66) For purine and quinoxaline, the B3LYP/aug-cc-pVTZ components from this work are used for the comparison since they match the experimental average polarizability reported by Shanker et al. Averaged experimental components reported by Shanker et al. are used for pyrimidine and pyrazine.

(67) Williams, G. J.; Stone, A. J. *Mol. Phys.* **2004**, *102*, 985–991.

(68) Oxtoby, D. W. *J. Chem. Phys.* **1980**, *72*, 5171–5176.

(69) Elcock, A. H.; Sept, D.; McCammon, J. A. *J. Phys. Chem. B* **2001**, *105*, 1504–1518.

(70) Miertus, S.; Scrocco, E.; Tomasi, J. *Chem. Phys.* **1981**, *55*, 117–129.

(71) Rowland, R. S.; Taylor, R. *J. Phys. Chem.* **1996**, *100*, 7384–7391.

CT800123C

JCTC Journal of Chemical Theory and Computation

Karplus Equation for $^3J_{HH}$ Spin−Spin Couplings with Unusual $^3J(180°) < {}^3J(0°)$ Relationship

R. H. Contreras,[†] R. Suardíaz,[‡] C. Pérez,[‡] R. Crespo-Otero,[‡] J. San Fabián,[§] and
J. M. García de la Vega*[,§]

Departamento de Física, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires and CONICET, Buenos Aires, Argentina, Departamento de Química Física, Facultad de Química, Universidad de la Habana, La Habana 10400, Cuba, and Departamento de Química Física Aplicada, Facultad de Ciencias, Universidad Autónoma de Madrid, 28049 Madrid, Spain

Received April 30, 2008

Abstract: Vicinal $^3J_{HH}$ coupling constants for monosubstituted ethane molecules present the unusual relationship $^3J_{HH}$ (180°) < $^3J_{HH}$ (0°) when the substituent contains bonding and antibonding orbitals with strong hyperconjugative interactions involving bond and antibond orbitals of the ethane fragment. This anomalous behavior is studied as a function of the substituent rotation for three model systems (propanal, thiopropanal, and 1-butene) at the B3LYP/TZVP level. The consistency of this level of theory to study this problem is previously established using different ab initio methods and larger basis sets. The origin of the unusual $^3J_{HH}(180°) - {}^3J_{HH}(0°)$ relationship is attributed to simultaneous σ/π hyperconjugative interactions $\sigma_{C_\alpha-H_\alpha} \rightarrow \pi^*_{C_c=X}$, and $\sigma_{C_\alpha-C_\beta} \rightarrow \pi^*_{C_c=X}$. These interactions depend on the substituent rotation and their effects are different for $^3J_{HH}(180°)$ than for $^3J_{HH}(0°)$. The modelization carried out shows an increase of those effects as the substituent changes from weaker ($CH{=}CH_2$) to stronger ($CH{=}S$) electron acceptor $\pi^*_{C=X}$.

1. Introduction

Vicinal NMR coupling constants were used extensively as stereochemical probes since Karplus pioneering works.[1,2] During the past decade, there was a renewed interest in vicinal spin−spin coupling constants (SSCC). Using these couplings as constraints in NMR structure refinement of proteins provides an important tool for increasing the definition of the peptide backbone and side chain conformations.[3]

In a previous work,[4] a valine dipeptide model was used, within the DFT framework, to obtain Karplus coefficients for $^3J_{XY}$ SSCCs whose X−C−C−Y dihedral angles are related to the dipeptide $\chi1$ angle. It is recalled that vicinal SSCCs can be represented by a Fourier series that reduces to the usual Karplus equation[1] if the SSCC asymmetry around $\phi = 180°$ and the higher cosine terms are neglected

$$^3J_{XY}(\varphi) = C_0 + C_1\cos(\varphi) + C_2\cos(2\varphi) \qquad (1)$$

Theoretically obtained SSCCs were compared with those inferred experimentally and, in general, they show a good agreement between them.[4] The largest differences were observed for $\chi1 = 0°$, where theoretical values were significantly larger than those obtained from empirical Karplus equations. Such theoretical couplings lead to an unusual positive coefficient C$_1$, eq 1, which can easily be related to the difference $^3J_{XY}(180°) - {}^3J_{XY}(0°) = -2C_1$ with $^3J_{XY}(180°) < {}^3J_{XY}(0°)$.[5] In the current literature, there are some experimental reports of positive C_1 coefficients. For instance, Chou et al.,[6] Lindorff-Larsen et al.,[7] and Juranić et al.[8] reported empirical positive C_1 coefficients for $^3J_{NC\gamma}$ in protein side chains. Positive C_1 coefficients were theoretically obtained by Case et al.[9] in valine, and by Chou et al.[6] in valine, threonine, and isoleucine. Also, $C_1 > 0$ values were reported for $^3J_{CH}$ in purine nucleotides by Munzarová et al.[10]

[*] Corresponding author e-mail: garcia.delavega@uam.es.

[†] Universidad de Buenos Aires and CONICET.

[‡] Universidad de la Habana.

[§] Universidad Autónoma de Madrid.

10.1021/ct800145h CCC: $40.75 © 2008 American Chemical Society
Published on Web 08/14/2008

All cases mentioned above in which $C_1 > 0$ present a molecular fragment **M** bonded to the coupling pathway of the $^3J_{HH}$ SSCC. This fragment **M** contains bonding and antibonding orbitals that can undergo strong hyperconjugative interactions with the coupling pathway fragment containing the coupling nuclei X−C−C−Y (X, Y = H, H; C, H; N, H; or N, C). In every case, the Fermi contact (FC) contribution determines the C_1 sign of the respective Karplus curves. Therefore, when intending to rationalize these facts, it is of primordial importance to pay attention to the transmission mechanism of the FC term.

The FC interaction originates when the electron density probability at the site of the coupling nuclei is not null. Several features of the FC transmission, including its angular dependence when the coupling nuclei are three or more bonds apart, have been well-known for many years.[11] In recent years, a deeper insight into how the FC term is transmitted through the electronic molecular structure was achieved,[12,13] and now it is known that its transmission is closely associated to the Fermi correlation, i.e., the "same-spin electron pair density", usually known as the "Fermi hole density". This indicates that departures from the classical Lewis structures by delocalization interactions should favor the transmission of long-range SSCC. Similarly, departures from the Lewis structures could affect notably all types of SSCCs dominated by the FC term. For this reason, in this work, special attention will be paid to departures from the Lewis structures. At present, the most frequently used approach to study these departures from the Lewis scheme is the natural bond orbitals (NBO) method of Weinhold et al.,[14] which gives a description of them and provides quantitative estimations of electron delocalization interactions. Usually, these delocalization interactions are classified as conjugative and hyperconjugative interactions.

In recent works, it was shown that σ-hyperconjugative interactions play a key role in transmitting long-range SSCCs in strained saturated compounds.[15] Also, it was observed that hyperconjugative interactions affect strongly one-,[16] two-,[17] and three-bond[18] SSCCs. Recently, it was reported[19] that strong hyperconjugative interactions between bonding and antibonding orbitals are relevant to the three bond contributions to $^{3,4}J_{CH}$ and those related to the carbonyl group in norbornanones, which can affect seriously both the three- and four-bond contributions.

In a previous work,[4] we demonstrated that the main contribution to the inversion of C_1 coefficient for $^3J_{H_\alpha H_\beta}$ in aminoacids is the C=O group, whereas the NH_2 group has a weak contribution. In this work, an interpretation of the "anomalous" behavior of the Karplus type eq 1 for $^3J_{HH}$ with $C_1 > 0$ is sought in terms of the molecular electronic structure. We select the propanal (Figure 1) as a simplified model to study the relationship between hyperconjugation interactions and coupling constants. Two additional models, thiopropanal and 1-butene (Figure 1) have been used to support the conclusions obtained in the propanal and to analyze the effect of different substituents (**M** = CH=S and CH=CH_2).

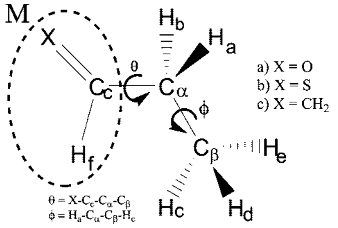


Figure 1. Propanal, thiopropanal, and 1-butene models.

2. Computational Details

Two kinds of geometries have been used in this paper. Initially, standard geometries[20] and tetrahedral angles were used for propanal to test the results of different methods and basis sets. Next, the geometries of propanal, thiopropanal and 1-butene were optimized at B3LYP/6−31G** level of theory which, is considered sufficiently accurate for the present purpose.[4,21] The geometries for the staggered (ϕ = 180°) and eclipsed (ϕ = 0°) conformers were calculated while the dihedral angle θ was varied in 30° steps from 0 to 330°. All degrees of freedom, except those of ϕ = 0° (in the eclipsed conformer) and θ, were optimized.

The 24 standard or partially optimized structures, 12 for $^3J_{HH}(180°)$ and 12 for $^3J_{HH}(0°)$, were used to calculate the four contributions to the $^3J_{HH}$: Fermi contact, spin dipolar (SD), paramagnetic spin−orbital (PSO) and diamagnetic spin−orbital (DSO). To test the quality of the results, we used the standard geometries to calculate the coupling constants of propanal with the following methods and basis set: B3LYP/TZVP, B3LYP/EPR-III, B3LYP/BS2, B3LYP/aug-cc-pVTZ-J, SOPPA/EPR-III, and CCSD(SOPPA)/EPR-III. The geometries partially optimized were used to calculate the coupling constants in propanal, thiopropanal and 1-butene at the B3LYP/TZVP level. In all studied cases, the FC term is by far the dominating one as can be appreciated with these examples for propanal (θ dihedral angle = 180°), $^3J_{HH}(180°)$ = 14.44 (FC = 15.11, SD = 0.02, PSO = 2.20, DSO = −2.89), and $^3J_{HH}(0°)$ = 13.36 (FC = 13.07, SD = 0.21, PSO = −0.18, DSO = 0.26), calculated at the B3LYP/TZVP level. NBO calculations have been performed at this same level of theory on those partial geometry optimizations. At this point, it is interesting to note that although individual hyperconjugative interactions depend on the basis set used to perform their calculations, trends of their angular dependences do not.

DFT calculations were performed with Gaussian03 software package,[22] SOPPA and CCSD(SOPPA) calculations were carried out with Dalton Software.[23] NBO 3.1[24] included in the Gaussian package[22] was used for the NBO calculations.

3. Results and Discussion

From the above considerations, it is reasonable to formulate a hypothesis connecting the peculiar features of $^3J_{HH}$ in Karplus-type equations satisfying the condition $^3J_{HH}(180°) < {}^3J_{HH}(0°)$ and hyperconjugative interactions taking place between bonding orbitals, or lone-pairs and antibonding

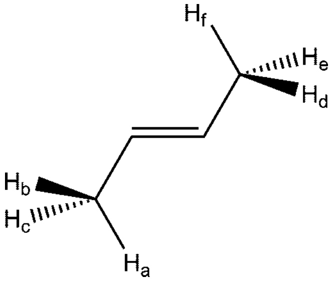


Figure 2. Homoallylic interproton coupling $J_{H_{b},H_{e}}$ whose coupling pathway is originated in the simultaneous existence of $\sigma_{C-H_b} \rightarrow \pi^*_{C=C}$ and $\sigma_{C-H_e} \rightarrow \pi^*_{C=C}$ σ/π hyperconjugative interactions.

orbitals of the molecular fragment **M** and those belonging to the $^3J_{HH}$ coupling pathway. In this way, propanal is taken as a model compound to elaborate a hypothesis about the influence of hyperconjugative interactions connecting the ethyl moiety with the carbonyl group (**M** fragment).

On the basis of the known[11] σ/π transmission of long-range homoallylic $^5J_{HH}$ in 2-butene, Figure 2, it can be expected that, for instance, the $^3J_{H_aH_c}$ SSCCs in propanal model of Figure 1 can be transmitted, in part, through the carbonyl π-electronic system by the simultaneous σ/π hyperconjugative interactions $\sigma_{C_\alpha-C_\beta} \rightarrow \pi^*_{C=O}$ and $\sigma_{C_\alpha-H_a} \rightarrow \pi^*_{C=O}$. It should be noted the close analogy between this assumption and the homoallylic coupling pathway shown in Figure 2. Any of these $\sigma_{C_\alpha-C_\beta} \rightarrow \pi^*_{C=O}$ and $\sigma_{C_\alpha-H_a} \rightarrow \pi^*_{C=O}$ interactions is zero whenever either the $\sigma_{C_\alpha-C_\beta}$ or $\sigma_{C_\alpha-H_a}$ bonds is contained in the carbonyl plane, i.e., for $\theta = 0$, 120, 180, and 300°. Similar effects are expected for the respective "opposite interactions", which for the sake of simplicity in this work will be called "back-donation" interactions, i.e., $\pi_{C=O} \rightarrow \sigma^*_{C_\alpha-C_\beta}$ and $\pi_{C=O} \rightarrow \sigma^*_{C_\alpha-H_a}$. It is noted that all these interactions follow either a $\sin^2 \theta$ or $\sin^2 \tau$ law, where θ and τ are the angles formed by the departure of the $\sigma_{C_\alpha-C_\beta}$ and $\sigma_{C_\alpha-H_a}$ bonds from the carbonyl plane, respectively. It is expected that the simultaneous occurrence of these two types of σ/π hyperconjugative interactions activates a second coupling pathway for both $^3J_{HH}(180°)$ and $^3J_{HH}(0°)$. However, the efficiency of such a σ/π coupling pathway could be different for each type of coupling and responsible for the sign inversion of C_1 coefficient.

Propanal Model. To select the functional and basis set appropriated for this work and to test the quality of the used calculations, the results from different methods and basis set are presented in Figure 3. The SSCCs for propanal have been calculated at the B3LYP functional and at the SOPPA and CCSD(SOPPA) approaches using EPR-III basis set. Although some qualitative differences are observed, for instance, the calculated values follow the relation $^3J_{HH}$ (CCSD(SOPPA)) < $^3J_{HH}$ (SOPPA) < $^3J_{HH}$ (B3LYP) independently of θ, in general, the qualitative dependence on θ is similar (see Figure 3). As regards to the differences $^3J_{HH}$ (180) − $^3J_{HH}$ (0), the ab initio methods give smaller values in magnitude but, again, the dependence on θ is similar for the three methods.

In addition, the B3LYP functional has been used with four basis sets of different size. The results for all basis sets are qualitatively similar. Moreover, the B3LYP results obtained with the smallest basis set (TZVP) yield the closest results to those of ab initio methods. This similitude between B3LYP/TZVP and ab initio/EPR-III could be attributed to a compensation between method and basis sets effects. On the other hand, B3LYP/TZVP has been recently tested successfully in the calculation of NMR coupling constants[25] and EPR hyperfine couplings.[21] Accordingly, we use B3LYP/TZVP to calculate $^3J_{HH}$ for the remaining systems presented in this work.

To quantify the hyperconjugative interaction and its effect on the SSCCs, we carried out $^3J_{HH}$ as well as NBO calculations for the propanal model. Using the notation shown in Figure 1, we calculated $^3J_{HH}(180°)$ and $^3J_{HH}(0°)$

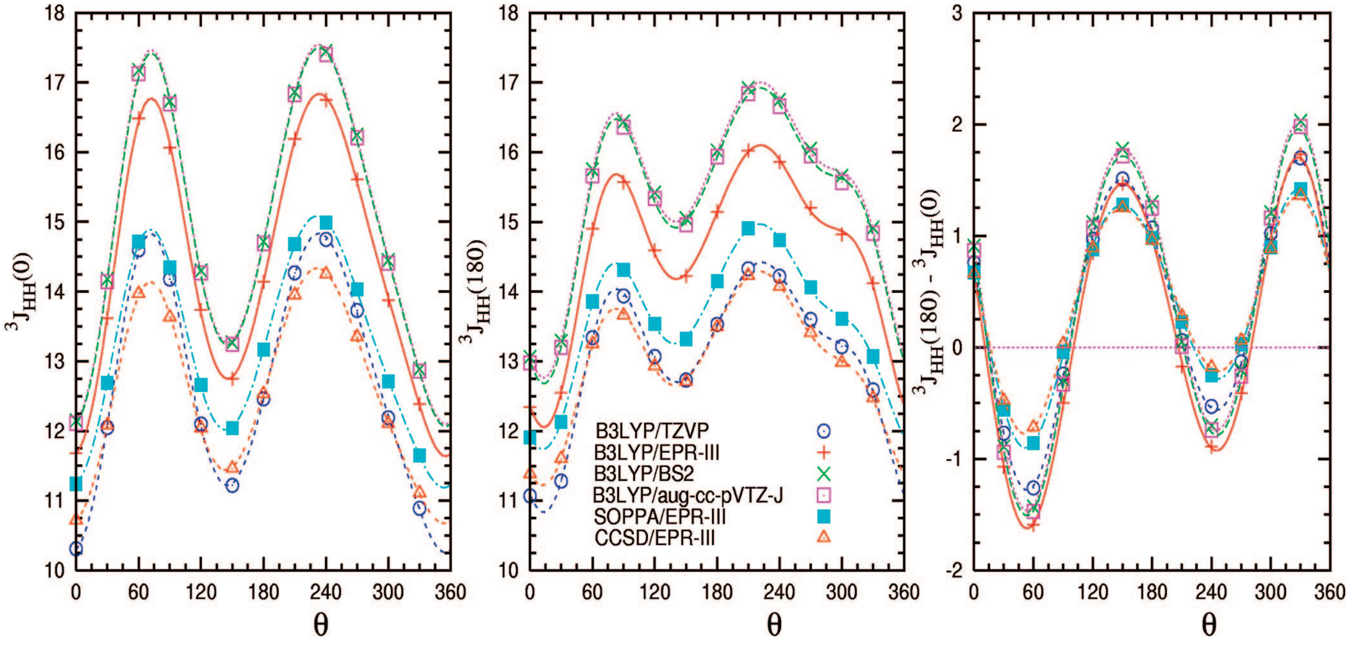


Figure 3. Ab initio and DFT calculated $^3J_{H_aH_c}(180°)$, $^3J_{H_aH_c}(0°)$, and $^3J_{H_aH_c}(180°) - ^3J_{H_aH_c}(0°)$ differences for propanal model.

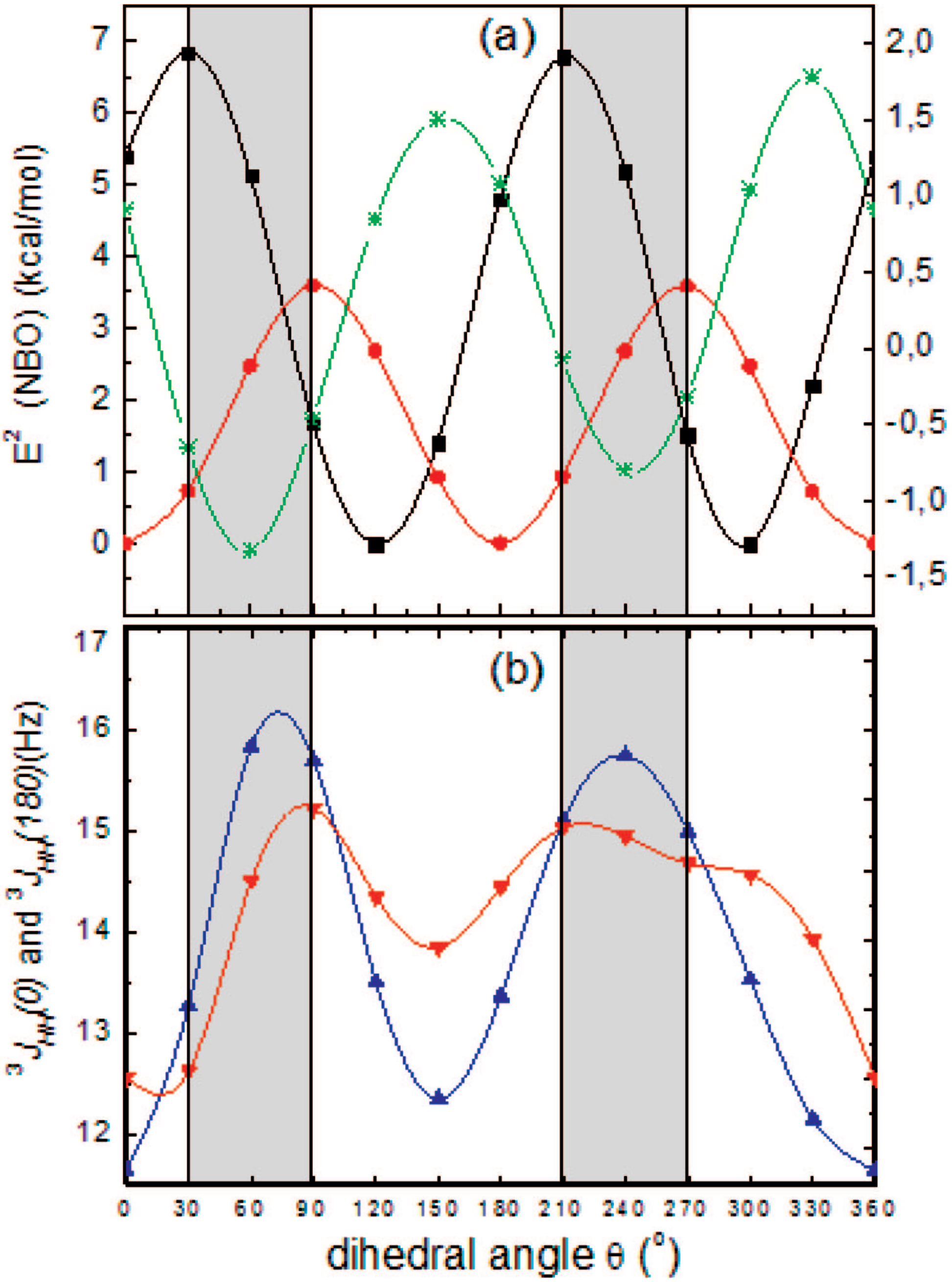


Figure 4. (a) Plots of E^2(NBO) $\sigma_{C_\alpha-C_\beta} \rightarrow \pi^*_{C=O}$ (•), $\sigma_{C_\alpha-H_a} \rightarrow \pi^*_{C=O}$ (■), and $^3J_{H_aH_c}(180°) - {}^3J_{H_aH_c}(0°)$ (*). (b) Plots of $^3J_{H_aH_c}(0°)$ (▲), and $^3J_{H_aH_c}(180°)$ (▼) vs dihedral angle θ. The zones of maxima hypercongujation (both, $\sigma_{C_\alpha-H_a} \rightarrow \pi^*_{C_c=O}$ and $\sigma_{C_\alpha-C_\beta} \rightarrow \pi^*_{C_c=O}$) that coincide with the zones where $^3J_{HH}(180) < {}^3J_{HH}(0)$ are shaded in these plots.

for the ethyl H$_a$Cα—CβH$_c$ moiety. Both types of couplings were calculated for different rigid rotations around the C$_c$—C$_\alpha$ bond, θ angle; values for $^3J_{H_aH_c}(180°)$ were obtained taking the staggered conformation of the ethyl group, while values of $^3J_{H_aH_c}(0°)$ were obtained taking the eclipsed conformation, Figure 1. Figure 4a shows the $^3J_{H_aH_c}(180°) - {}^3J_{H_aH_c}(0°)$ difference as well as the hyperconjugative interactions $\sigma_{C_\alpha-H_a} \rightarrow \pi^*_{C_c=O}$ and $\sigma_{C_\alpha-C_\beta} \rightarrow \pi^*_{C_c=O}$ versus θ. The following features of these three plots shown in Figure 4a are worth highlighting:

(a) The values of $^3J_{HH}(0°)$ and $^3J_{HH}(180°)$ depend notably on θ, this dependence being nonsymmetric around $\theta = 180°$. This asymmetry seems to originate mainly on the proximity between the carbonyl O and the H$_a$ atom for θ =120° and the carbonyl O and the H$_c$ atom for θ = 300° (the former for both the staggered and the eclipsed conformation, and the latter only for the ethyl staggered conformation).

(b) Both σ/π hyperconjugative interactions follow the well-known $\sin^2 \theta$ dependence, lagging the $\sigma_{C_\alpha-C_\beta} \rightarrow \pi^*_{C_c=O}$ plot 60° with respect to that of $\sigma_{C_\alpha-H_a} \rightarrow \pi^*_{C_c=O}$.

As shown in Figure 4b, both types of couplings depend on θ angle, the sensitivity of $^3J_{HH}(0°)$ being higher than that of $^3J_{HH}(180°)$. This indicates that the “activated” coupling pathway due to the hyperconjugative interactions involving the carbonyl group shows a different efficiency for $^3J_{HH}(180°)$ and $^3J_{HH}(0°)$. In Figure 4, the zones of maxima hyperconjugation (both, $\sigma_{C_\alpha-H_a} \rightarrow \pi^*_{C_c=O}$ and $\sigma_{C_\alpha-C_\beta} \rightarrow \pi^*_{C_c=O}$), that coincide with the zones where $^3J_{HH}(180) < {}^3J_{HH}(0)$, are

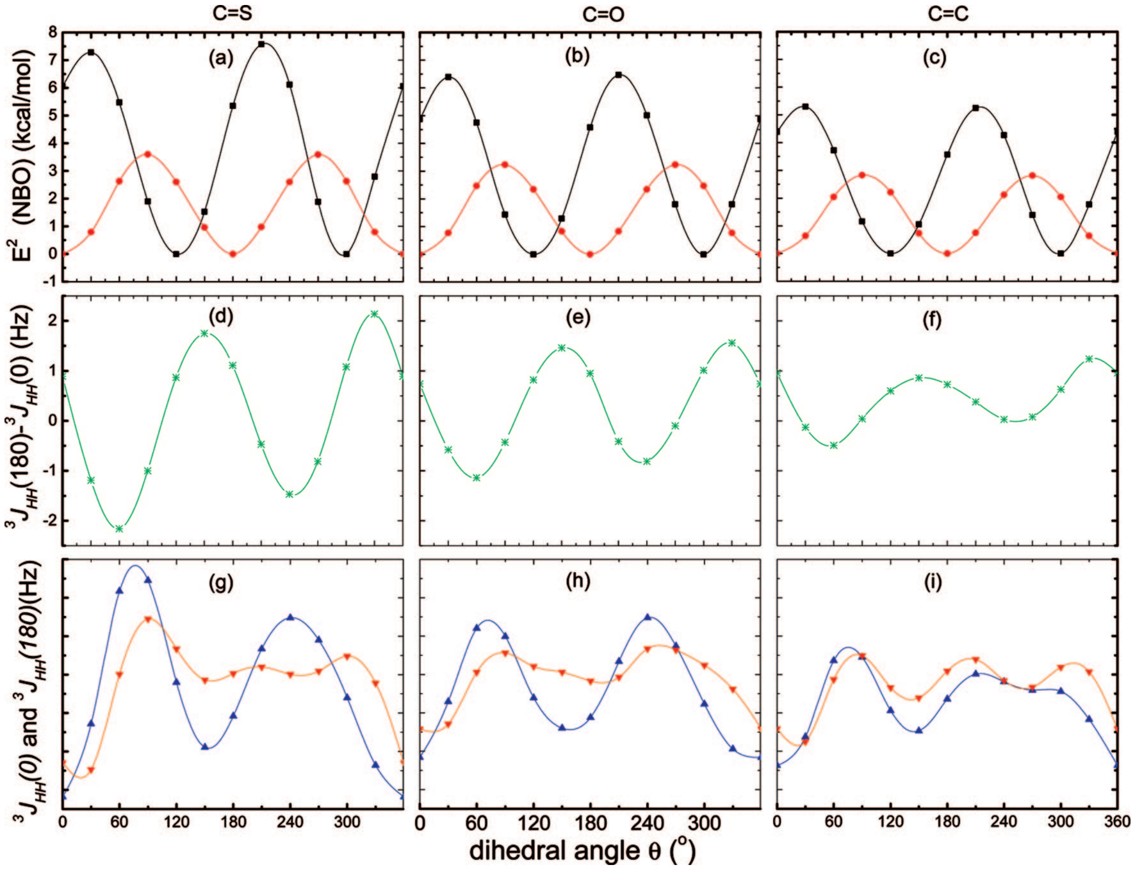


Figure 5. Plots of E^2(NBO) $\sigma_{C_\alpha-C_\beta} \rightarrow \pi^*_{C=X}$ (•) and $\sigma_{C_\alpha-H_a} \rightarrow \pi^*_{C=X}$ (■) (first row); $^3J_{H_aH_c}(180°) - {}^3J_{H_aH_c}(0°)$ (second row); and plots of $^3J_{H_aH_c}(0°)$ (▲) and $^3J_{H_aH_c}(180°)$ (▼) (third row) vs dihedral angle θ, for thiopropanal (first column), propanal (second column), and 1-butene (third column).

highlighted. Those zones correspond to an inversion of the C_1 coefficient.

For $\theta = 0$, 120, 180, and 300°, the behavior of both $^3J_{HH}(0)$ and $^3J_{HH}(180)$ SSCCs cannot be described by the σ/π hyperconjugative interactions because either $\sigma_{C_\alpha-C_\beta} \rightarrow \pi^*_{C_c=O}$ or $\sigma_{C_\alpha-H_a} \rightarrow \pi^*_{C_c=O}$ interactions are equal to zero. For these four conformers, $^3J_{HH}(180°)$ is notably larger than in $^3J_{HH}(0°)$ (about 1 Hz). It is interesting to remark that in ethane molecule, this difference, $^3J_{HH}(180) - {}^3J_{HH}(0)$, calculated at the MCSCF/RAS level, is around 0.94 Hz,[26] which is similar to that found above.

For $\theta = 240°$ the σ/π hyperconjugative interactions seem to be the main contribution to both SSCCs in defining the $^3J_{HH}(180°) - {}^3J_{HH}(0°)$ value. On the other hand, for $\theta = 270°$ the carbonyl O atom is nearing the $\sigma_{C_\beta-H_c}$ bond, the nearest approach being for $\theta = 300°$. It is observed in Figure 4b that $^3J_{H_aH_c}(180°)$ for $\theta = 300°$ is similar to $^3J_{H_aH_c}(180°)$ for $\theta = 120°$, yielding the expected conclusion that the proximity effect on $^3J_{HH}(180°)$ is similar whether either the $\sigma_{C_\alpha-H_a}$ or $\sigma_{C_\beta-H_c}$ bonds are close to the carbonyl oxygen.

It is important to remark that the coupling pathway defined by the simultaneous $\sigma_{C_\alpha-H_a} \rightarrow \pi^*_{C_c=O}$, $\sigma_{C_\alpha-C_\beta} \rightarrow \pi^*_{C_c=O}$ hyperconjugative interactions and their respective back-donations refers only to the FC term. The sum of the SD, PSO, and DSO contributions are larger in absolute value for $^3J_{HH}(180°)$ than for $^3J_{HH}(0°)$; however, the SD + PSO + DSO sum for each of them is notably insensitive to the θ angle and therefore the $^3J_{HH}(180°) - {}^3J_{HH}(0°)$ trend is by far defined by the FC term.

Thiopropanal and 1-Butene Models. The analysis presented above for propanal supports the hypothesis about the electronic origin of the $^3J_{HH}(180°) < {}^3J_{HH}(0°)$ relationship in this model system. Because hyperconjugative interactions of types $\sigma_{C_\alpha-H_a} \rightarrow \pi^*_{C_c=O}$, $\sigma_{C_\alpha-C\beta} \rightarrow \pi^*_{C_c=O}$, and their respective back-donations play the main influence on such a relationship, it is considered convenient to look for similar models where the main difference with propanal is due to σ/π interactions. For instance, if the carbonyl group in propanal is replaced by a thiocarbonyl group, Figure 1, then it is expected that the $\pi^*_{S=C}$ antibonding orbital to be a better electron acceptor than the $\pi^*_{O=C}$ antibonding orbital. Hence, in thiopropanal, the relevant σ/π interactions should be more important than in propanal. On the other hand, if in propanal the carbonyl group is replaced by a vinyl group, Figure 1, then the relevant σ/π interactions in 1-butene should be weaker than in propanal because the $\pi^*_{C=C}$ antibonding orbital is a poorer electron acceptor than both carbonyl and thiocarbonyl group.

Figure 5a−c shows plots of the $\sigma_{C_\alpha-H_a} \rightarrow \pi^*_{C_c=X}$, and $\sigma_{C_\alpha-C_\beta} \rightarrow \pi^*_{C_c=X}$ interactions vs θ for thiopropanal (X = S), propanal (X = O), and 1-butene (X = CH$_2$), respectively. These interactions follow the expected trend, i.e., they decrease along the series from thiopropanal to 1-butene, whereas the angular dependence is similar in all three cases.

The $^3J_{HH}(180°) - {}^3J_{HH}(0°)$ differences for the three model molecules are plotted in Figure 5d−f. They follow qualitatively similar trends and their amplitude decreases notably along this series, paralleling the behavior of σ/π interactions displayed in Figures 5a−c. These results strongly support the hypothesis about the importance played by σ/π interactions in defining the $^3J_{HH}(180°) < {}^3J_{HH}(0°)$ relationship. The plots of $^3J_{HH}(0°)$ and $^3J_{HH}(180°)$ vs θ are shown in Figures 5g−i. Both plots show essentially similar behavior along this series, where differences can be easily rationalized as originating in the two effects discussed above for propanal, i.e., the σ/π hyperconjugative interactions and the proximity effect between the carbonyl and ethyl groups. Along this series, it is interesting to note that the different efficiency of the "activated" coupling pathway for $^3J_{HH}(0°)$ and $^3J_{HH}(180°)$ increases when increasing the σ/π interactions, reinforcing the possibility of obtaining $^3J_{HH}(180°) < {}^3J_{HH}(0°)$ when the molecular fragment **M** shows stronger hyperconjugative interactions with the ethyl group.

4. Conclusions

The origin of the unusual C_1 positive Fourier coefficient, or the anomalous $^3J_{HH}(180°) - {}^3J_{HH}(0°)$ relationship found in some molecules has been studied at the DFT level in three model molecules.

The effects of the substituent and its rotation on $^3J_{HH}(180°)$ and $^3J_{HH}(0°)$ SSCC have been investigated at B3LYP level and with the TZVP basis set after checking that these results are similar to those of ab initio levels (SOPPA and CCSD-(SOPPA)) and to those of larger basis sets (EPR-III, aug-cc-pVTZ, and BS2).

It is concluded that strong hyperconjugative interactions involving both bonding and antibonding orbitals of the coupling pathway as well as of the carbonyl group are essential to explain the "anomalous" behavior of $^3J_{HH}$. Such interactions define an additional coupling pathway for $^3J_{HH}$. For propanal, this additional pathway partially can be assigned to simultaneous σ/π hyperconjugative interactions $\sigma_{C_\alpha-H_a} \rightarrow \pi^*_{C_c=X}$, and $\sigma_{C_\alpha-C_\beta} \rightarrow \pi^*_{C_c=X}$ (X = O). These interactions are more efficient for $^3J_{HH}(0°)$ than for $^3J_{HH}(180°)$. Moreover, for θ angles where these interactions are strongest, the $^3J_{HH}(0°)$ couplings are larger than $^3J_{HH}(180°)$ and the C_1 coefficient becomes positive.

Analogous interactions have been detected in thiopropanal (X = S) and 1-butene (X = CH_2). In the former, the hyperconjugative effects are stronger than in propanal because of a better electron acceptor behavior of the $\pi^*_{S=C}$ antibonding orbital, whereas in butane, the weaker effect is attributed to the poorer electron acceptor behavior of the $\pi^*_{C=C}$ antibonding orbital.

Acknowledgment. R.H.C. gratefully acknowledges economic support from UBACYT (X222) and CONICET (PIP 5119/05). R.S. and R.C.O. acknowledge a research fellowship from Universidad Autónoma de Madrid. J.M.G.V. and J.S.F. gratefully acknowledge the financial support from the Dirección General de Enseñanza Superior e Investigación Científica of Spain (DGESIC, projects CTQ2005-04469, CTQ2007-63332, and CTQ2007-66547). Computer time has been partially provided by the Centro de Computación Científica of Universidad Autónoma de Madrid.

References

(1) (a) Karplus, M. *J. Chem. Phys.* **1959**, *30*, 11. (b) Karplus, M. *J. Phys. Chem.* **1960**, 1793. (c) Karplus, M. *J. Am. Chem. Soc.* **1963**, *85*, 2870.

(2) (a) Bystrov, V. F. *Prog. NMR Spectrosc.* **1976**, *10*, 41. (b) Haasnoot, C. A. G.; De Leeuw, F. A. A. M.; Altona, C. *Tetrahedron* **1980**, *36*, 2783. (c) Imai, K.; Osawa, E. *Magn. Reson. Chem.* **1990**, *28*, 668. (d) Altona, C.; Ippel, J. H.; Hoekzema, A. J. A. W.; Erkelens, C.; Groesbeek, M.; Donders, L. A. *Magn. Reson. Chem.* **1989**, *27*, 564. (e) Díez, E.; San Fabián, J.; Guilleme, J.; Altona, C.; Donders, L. A. *Mol. Phys.* **1989**, *68*, 49.

(3) Rule, G. S. Hitchens, T. K. Protein Structure Determination. In *Fundamentals of Protein NMR Spectroscopy*; Springer: Dordrecht, The Netherlands, 2006; p 387.

(4) Suardíaz, R.; Pérez, C.; García de la Vega, J. M.; San Fabián, J.; Contreras, R. H. *Chem. Phys. Lett.* **2007**, *442*, 119.

(5) Pérez, C.; Löhr, F.; Rüterjans, H.; Schmidt, J. M. *J. Am. Chem. Soc.* **2001**, *123*, 7081.

(6) Chou, J. J.; Case, D. A.; Bax, A. *J. Am. Chem. Soc.* **2003**, *125*, 8959.

(7) Lindorff-Larsen, K.; Best, R. B.; Vendruscolo, M. *J. Biomol. NMR* **2005**, *32*, 273.

(8) Juranic, N.; Atanasova, E.; Moncrieffe, M. C.; Prendergast, F. G.; Macura, S. *J. Magn. Reson.* **2005**, *175*, 222.

(9) Case, D. A.; Scheurer, C.; Brüschweiller, R. *J. Am. Chem. Soc.* **2000**, *122*, 10390.

(10) (a) Munzarová, M. L.; Sklenár, V. *J. Am. Chem. Soc.* **2002**, *124*, 10666. (b) Munzarová, M. L.; Sklenár, V. *J. Am. Chem. Soc.* **2003**, *125*, 3649.

(11) Barfield, M.; Chakrabarti, B. *Chem. Rev.* **1969**, *69*, 757.

(12) Castillo, N.; Matta, C. F.; Boyd, R. J. *J. Chem. Inf. Model.* **2005**, *45*, 354.

(13) Soncini, A.; Lazzeretti, P. *J. Chem. Phys.* **2003**, *119*, 1343.

(14) (a) Reed, A. E.; Curtiss, L. A.; Weinhold, F. *Chem. Rev.* **1988**, *88*, 899. (b) Weinhold, F. Natural Bond Orbital Methods. In *Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P. A., Schaefer, H. F., III, Schreiner, P. R., Eds.; John Wiley & Sons: Chichester, U.K., 1998; Vol. 3, p 1792.

(15) Contreras, R. H.; Esteban, A. L.; Díez, E.; Head, N. J.; Della, E. W. *Mol. Phys.* **2006**, *104*, 485.

(16) Tormena, C. F.; Rittner, R.; Contreras, R. H.; Peralta, J. E. *J. Phys. Chem. A* **2004**, *108*, 7762.

(17) Contreras, R. H.; Esteban, A. L.; Díez, E.; Della, E. W.; Lochert, I. J.; dos Santos, F. P.; Tormena, C. F. *J. Phys. Chem. A* **2006**, *110*, 4266.

(18) Esteban, A. L.; Galache, M. P.; Mora, F.; Díez, E.; Casanueva, J.; San Fabián, J.; Barone, V.; Peralta, J. E.; Contreras, R. H. *J. Phys. Chem. A* **2001**, *105*, 5298.

(19) dos Santos, F. P.; Tormena, C. F.; Contreras, R. H.; Rittner, R.; Magalhaes, A. *Magn. Reson. Chem.* **2008**, *46*, 107.

(20) Pople, J. A.; Gordon, M. *J. Am. Chem. Soc.* **1967**, *89*, 4253.

(21) Hermosilla, L.; Calle, P.; García de la Vega, J. M.; Sieiro, C. *J. Phys. Chem. A* **2005**, *109*, 1114.

(22) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez C.; Pople J. A. *Gaussian 03, Revision D.01*; Gaussian, Inc.: Wallingford, CT, 2004.

(23) Helgaker, T.; Aa. Jensen, H. J.; Jorgensen, P.; Olsen, J.; Ruud, K.; Agreni, H.; Andersen, T.; Bak, K. L.; Bakken, V.; Christiansen, O.; Dahle, P.; Dalskov, E. K.; Enevoldsen, T.; Fernandez, B.; Heibergi, H.; Hettema, H.: Jonsson, D.; Kirpekar, S.; Kobayashi, R.; Koch, H.; Mikkelsen, K. V.; Norman, P.; Packer, M. J.; Saue, T.; Taylor, P. R.; Vahtras, O. Dalton, An Electronic Structure Program, releases 1.2; University of Oslo: Oslo, Norway, 1997.

(24) Glendening, E. D.; Reed, A. E.; Carpenter, J. E.; Weinhold, F. *Gaussian NBO, version 3.1*; Gaussian Inc.: Pittsburgh, PA.

(25) Suardíaz, R.; Pérez, C.; Crespo-Otero.; García de la Vega, J. M.; San Fabián, J. *J. Chem. Theory Comput.* **2008**, *4*, 448.

(26) Guilleme, J.; San Fabián., J.; Casanueva, J.; Díez, E. *Chem. Phys. Lett.* **1999**, *314*, 168.

CT800145H

JCTC Journal of Chemical Theory and Computation

Performance of the Density Functional Theory/ Multireference Configuration Interaction Method on Electronic Excitation of Extended π-Systems

Christel M. Marian* and Natalie Gilka[†]

Institute of Theoretical and Computational Chemistry, Heinrich-Heine-University Düsseldorf, Universitätsstraße 1, 40225 Düsseldorf, Germany

Received May 19, 2008

Abstract: The combined density functional theory/multireference configuration interaction (DFT/MRCI) method [Grimme and Waletzke. *J. Chem. Phys.* 1999, *111*, 5645] has been employed to study the ^1L_a and ^1L_b states of linear polyacenes and the low-lying triplet and singlet states of linear polyenes and diphenyl-polyenes. We have systematically investigated the dependence of the electronic state properties on technical parameters of the calculations such as the atomic orbital basis set or the geometry optimization approach. The choice of basis set appears to be of minor importance whereas the excitation energies of the polyenes are quite sensitive to the ground-state geometry parameters. The DFT/MRCI energies at the B3-LYP optimized geometries systematically underestimate the experimental values, but we do not observe a bias toward one or the other type of state. The energy gaps between the electronically excited states are reproduced very well. In particular, this applies also to the first excited singlet 2 $^1A_g^-$ and the optically bright $^1B_u^+$ state of the polyenes. The latter appears to be the S_3 or even S_4 state in longer polyenes where the multiconfigurational $^1B_u^-$ state represents S_2. Frequencies and intensities of the excited-state absorption from the 2 $^1A_g^-$ state are found to be strongly geometry dependent.

1. Introduction

Many biologically relevant pigments such as carotenoids and retinal contain extended polyene chromophores. Naturally occurring carotenoids possess between 7 and 11 conjugated double bonds and exhibit a variety of low-lying electronic states, many of which are not easily accessed spectroscopically because the corresponding one-photon transition from the ground state is forbidden (*gerade*-states, triplets, double excitations).[1] A reliable quantum chemical description of these states could therefore help understanding the intricate photophysical relaxation mechanisms these molecules undergo after electronic excitation.

In recent years, time-dependent density functional theory (TDDFT) has emerged as standard tool for the evaluation of electronically excited states of large molecules for which traditional wave function based methods are not feasible. Unfortunately, TDDFT in combination with standard functionals fails in describing the correct ordering of the two lowest excited singlet states of linear polyenes and carotenoids.[2−5] The Tamm−Dancoff approximation (TDA) to TDDFT appears to perform better in this respect.[2] However, the success of this method was recently shown to be based on fortuitous cancelation of errors.[6] It has been firmly established that the optically bright ^1B_u state which originates from the (HOMO → LUMO) single excitation is not the lowest excited singlet state in linear polyenes, possibly except for butadiene.[7,8] Instead, the S_1 state possesses ^1A_g symmetry and exhibits a highly multiconfigurational character with the (HOMO → LUMO)^2, (HOMO − 1 → LUMO), and (HOMO → LUMO + 1) configurations as leading terms.[9−11] Complete active space second-order perturbation theory (CASPT2)

* To whom correspondence should be addressed. Tel.: +49-211-8113209. Fax: +49-211-8113446. E-mail: Christel.Marian@uni-duesseldorf.de.

[†] Present address: Department of Chemistry, University of Warwick, CV4 7AL, Coventry, UK.

10.1021/ct8001738 CCC: $40.75
Published on Web 08/15/2008

or more general multireference second-order perturbation theory (MRMP2) treatments of the electron correlation are capable of describing these states and their energetic ordering properly.[12−14] With increasing chain length it becomes more and more demanding, however, to include the appropriate orbitals in the active space. To the best of the authors' knowledge, no CASPT2 or MRMP2 treatment has been published for polyenes with seven or more conjugated double bonds. Coupled-cluster calculations with single and double excitations (albeit with an approximate treatment of the doubles, CC2) can in principle be applied to study the electronically excited states of large molecules. However, because of the double-excitation character of the 2 ^1A_g state, also this method places the 2 ^1A_g state considerably above the 1 ^1B_u state in hexatriene[15−17] and octatetraene.[17] The second-order algebraic diagrammatic construction (ADC(2)) method[18] allows in principle for the consistent treatment of doubly excited states. Applications to several polyenes including a treatment of doubly excited configurations through zeroth (ADC(2)-s) or first order (ADC(2)-x) are available[6] and will be discussed below.

Another spectroscopically important class of molecules with extended π-systems are linear condensed acenes. Parac and Grimme[19] could show that TDDFT yields substantial errors in the description of the short-axis polarized L_a state while the long-axis polarized L_b is described well. The CC2 method, on the other hand, achieves a balanced description of the two states, in this case. Here, the difficulties do not arise from a double excitation character as for the 2 ^1A_g state of the polyenes. For benzene they were shown to result from large dynamical $\sigma-\pi$ polarization effects.[20]

In the present work, we investigate the performance of the combined density functional theory and multireference configuration interaction method (DFT/MRCI)[21] on low-lying singlet and triplet states of polyenes, α,ω-diphenyl-polyenes, and polyacenes. The DFT/MRCI method has been shown to yield reliable excitation energies and transition moments at reasonable cost for a variety of molecules.[17,21−27] Particular emphasis is put on the above-mentioned critical cases, namely the 2 ^1A_g and 1 ^1B_u states of polyenes and α,ω-diphenyl-polyenes as well as the L_a and L_b states of polyacenes. In addition, we study trends of the properties of further low-lying states in the series of *all-trans*-polyenes beginning with 1,3,5-hexatriene and extending to 1,3,5,7,9,11,13,15,17,19,21,23,25-hexacosatridecaene, as the location of these states may be important for the excited-state absorption (ESA) and for the relaxation dynamics following the population of the optically bright (HOMO → LUMO) singlet excited state.

2. Methods and Technical Details

Computations on larger polyenes may become cumbersome if extended basis sets are used. Therefore, our first issue was a search for technical parameters of the calculation with optimal cost/performance ratio.

Three qualitatively different basis sets from the Turbomole library[28] were employed: the split valence basis set with (d) polarization functions for non-hydrogen atoms, (SV(P)), the valence triple-ζ basis set with polarization functions (d,p) (TZVP), and the valence triple-ζ basis set with a double set of polarization functions (2d1f,2p1d) (TZVPP). If not specified otherwise, the equilibrium nuclear arrangements of the electronic ground states were determined using density functional theory (DFT) in combination with a restricted closed-shell Kohn−Sham determinant. The geometries of the first excited triplet states were optimized utilizing unrestricted DFT (UDFT). Finally, TDDFT[29] was used to obtain the minimum geometries of the singlet-coupled (HOMO → LUMO) excited states. These calculations were carried out employing the Turbomole suite of programs.[30] We also tested the performance of different density functionals for geometry optimization. Among these, the local B-LYP functional[31,32] is the less expensive one in terms of computer time because one can make use of the resolution-of-the-identity (RI) approximation.[33,34] The second functional used is the well-known B3-LYP functional.[32,35,36] It typically yields reliable bond distances and frequencies. Since it is a hybrid functional which includes 20% Hartree−Fock exchange, no use of the RI approximation can be made. The third functional, BH-LYP,[32,37] includes 50% Hartree−Fock exchange. In combination with the DFT/MRCI approach it is the standard functional used for generating the MO basis and Fock matrix elements. With regard to minimum geometries it is known to yield typically somewhat too compact molecular structures. As purely wave function based methods Hartree−Fock (HF) and second-order Møller−Plesset perturbation theory (MP2) are employed in a few test cases.

Electronic excitation energies are evaluated by means of the DFT/MRCI method.[21] The idea behind this approach is to include major parts of dynamic electron correlation by density functional theory whereas short to medium-sized MRCI expansions take account of static correlation effects. In this way, severe size-extensivity problems can be avoided even for systems with many valence electrons. The configurations in the MRCI expansion are built up from Kohn−Sham (KS) orbitals of a closed-shell reference state. In the effective DFT/MRCI Hamiltonian, five empirical parameters (scaling factors for Coulomb and exchange integrals as well as energy cutoff parameters) are employed that depend only on the multiplicity of the desired state, the number of open shells of a configuration, and the type of density functional employed, but not on the specific atom or molecule. To avoid double-counting of dynamic correlation, the MRCI expansion is kept short by extensive configuration selection. Currently, optimized parameter sets for the effective DFT/MRCI Hamiltonian are available in combination with the BH-LYP functional. We employ the original set of parameters[21] here in combination with an orbital selection threshold of $1.0E_\mathrm{H}$. For details concerning the integration of DFT information into the MRCI procedure and the parameter fitting, we refer to the original publications.[21]

3. Results and Discussion

3.1. Benchmark Systems. *3.1.1. Linear Conjugated Acenes.* TDDFT in combination with standard functionals has been shown to give dramatic failures for the short-axis polarized ^1L_a state of linear conjugated acenes.[19] The ^1L_a

Table 1. Calculated Vertical Absorption Energies ΔE_{vert} [eV] of Linear Condensed Acenes in Comparison with Previous Theoretical Results and Experimental Values[a]

	DFT/MRCI					
number of rings	SV(P)	TZVP	BP86[b]	B3-LYP[c]	CC2[d]	exp[e]
		1 $^1B_{2u}$, ^1L_a state				
2	4.70 (0.1257)	4.66 (0.1222)	4.11	4.38	4.88	4.66
3	3.53 (0.1279)	3.51 (0.1249)	2.95	3.21	3.69	3.60
4	2.75 (0.1111)	2.74 (0.1088)	2.17	2.43	2.90	2.88
5	2.22 (0.0929)	2.22 (0.0916)	1.63	1.89	2.35	2.37
6	1.86 (0.0749)	1.85 (0.0744)	1.23	1.49	1.95	2.02
8	1.46 (0.0407)	1.44 (0.0418)	0.68	0.94	1.43	1.58
		1 $^1B_{3u}$, ^1L_b state				
2	4.13 (0.0002)	4.15 (0.0001)	4.13	4.26	4.47	4.46
3	3.56 (0.0012)	3.59 (0.0007)	3.64	3.87	3.89	3.64
4	3.20 (0.0034)	3.22 (0.0023)	3.24	3.47	3.52	3.39
5	2.96 (0.0069)	2.99 (0.0054)	2.96	3.21	3.27	3.12
6[f]	2.72 (0.0051)	2.76 (0.0051)	2.76	3.02	3.09	2.87
	2.90 (0.0066)	2.93 (0.0045)				
8	2.77 (0.0269)	2.78 (0.0244)	2.50	2.77	2.87	

[a] Oscillator strengths $f(r)$ are given in parentheses. [b] See ref 19 for TDDFT(BP86) using Dunning's cc-pVTZ basis at the ground-state geometry obtained from DFT(B3-LYP) calculations in a TZVP basis. [c] See ref 19 for TDDFT(B3-LYP) using Dunning's cc-pVTZ basis at the ground-state geometry obtained from DFT(B3-LYP) calculations in a TZVP basis. [d] See ref 19 for CC2 using Dunning's cc-pVTZ basis at the ground-state geometry obtained from DFT(B3-LYP) calculations in a TZVP basis. [e] Derived from 0−0 transition energies in solution.[38] For details see ref 19. [f] The (HOMO − 2 → LUMO) and (HOMO → LUMO + 2) configurations are spread over the 1 $^1B_{3u}$ and 2 $^1B_{3u}$ DFT/MRCI wave functions. Due to their energetic proximity and mixed wave function character, the two low-lying $^1B_{3u}$ states exhibit similar oscillator strengths and together make up the ^1L_b state in this molecule. For this reason, two values are listed.

state has $^1B_{2u}$ symmetry and results from the (HOMO → LUMO) excitation. In a valence-bond picture, this state exhibits large contributions from ionic components. In addition to the ^1L_a state, Parac and Grimme[19] studied the long-axis polarized ^1L_b state. The latter state exhibits $^1B_{3u}$ symmetry. Its electronic structure is mainly covalent. For molecules with 2−4 conjugated rings, its wave function is dominated by a nearly equal mixture of (HOMO − 1 → LUMO) and (HOMO → LUMO + 1) excitations.

We computed vertical DFT/MRCI excitation energies at DFT(B3-LYP) optimized geometries for polyacenes with 2−6 and with 8 rings using both SV(P) and TZVP basis sets. Our results are collected in Table 1 and are compared to the TDDFT and CC2 excitation energies of Parac and Grimme.[19] These authors also published estimates of vertical experimental excitation energies which they derived from experimental 0−0 data[38] by correcting for solvent and relaxation effects. These estimates are displayed in Table 1 as well. It is seen that the DFT/MRCI results do not suffer from the flaws of the TDDFT treatments. Actually, their accuracy is comparable to the one of the ab initio CC2 method, but with a tendency of slightly underestimating the ^1L_a and ^1L_b excitation energies whereas CC2 overestimates the experimental values by approximately the same amount. With regard to the treatment of large π-systems, it is encouraging that the reduction of the basis set quality from TZVP to SV(P) has almost no effect on the DFT/MRCI excitation energies. The underlying reason for this behavior is the fact that the (virtual) DFT orbitals are significantly less diffuse than HF orbitals. Thus, the basis set size does not influence the orbital energies as much as in HF.

Before we take a more detailed look at the ^1L_b and ^1L_a energies, a short overview over the energetic order of molecular orbitals will be given. In the compounds with even number of rings, the HOMO − 1 belongs to the b_{1u} irreducible representation (irrep) of the D_{2h} molecular point group, the HOMO has a_u symmetry, while the virtual orbitals LUMO and LUMO + 1 transform according to the b_{2g} and b_{3g} irreps, respectively. In anthracene, the order is reversed with b_{2g} and b_{3g} orbitals building HOMO − 1 and HOMO, respectively, and LUMO and LUMO + 1 belonging to the b_{1u} and a_u irreps, respectively. According to our calculations, the order of the frontier orbitals becomes irregular for the polyacenes with more than 4 rings. While the identity of HOMO and LUMO is preserved, we find the highest occupied b_{2g} orbital as HOMO − 2 in pentacene, with an orbital of a_u symmetry forming HOMO − 1. A similar situation is found for the unoccupied orbitals. Here, an orbital of b_{2g} symmetry forms the LUMO + 1 while the lowest unoccupied orbital of a_u symmetry is LUMO + 2. Accordingly, we find the (HOMO − 2 → LUMO) and (HOMO → LUMO + 2) excitations to be dominating the electronic structure of the ^1L_b state. Similarly, the highest occupied b_{1u} orbital has been shifted to HOMO − 2 in hexacene and the lowest unoccupied b_{3g} orbital is LUMO + 2 here. HOMO − 1 exhibits b_{3g} symmetry while the LUMO + 1 belongs to the b_{1u} irrep. The tendency of shifting the orbitals, involved in the ^1L_b excitation, away from the Fermi level continues in octacene. Here, the corresponding b_{1u} and b_{3g} orbitals yield HOMO − 3 and LUMO + 3, respectively.

A close look at the excitation energies of particular members of this series in Table 1 shows that the DFT/MRCI and CC2 methods correctly predict the swap of energetic order of the ^1L_b and ^1L_a states to occur between $n = 2$ and $n = 3$. In polyacenes with 3−6 conjugated rings, the ^1L_a state represents the S_1 state. In octacene, our DFT/MRCI calculation yields 2 ^1A_g as the first excited singlet state. Its MRCI expansion is dominated by double excitations with the leading term being the (HOMO → LUMO)^2 configuration. In contrast to the situation in linear polyenes (see below)

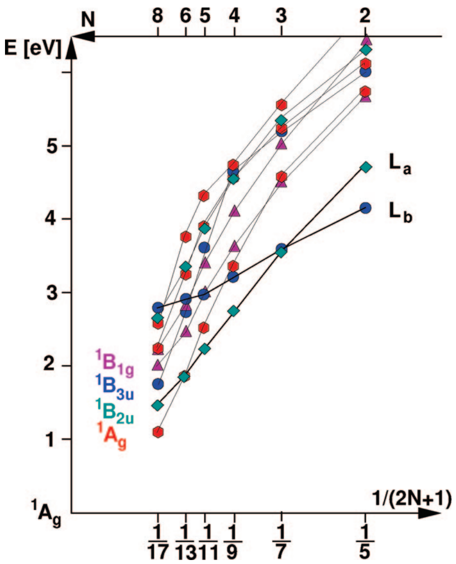


Figure 1. DFT/MRCI vertical electronic excitation energies of polyacenes at the respective 1 ^1A_g ground-state minimum geometry as functions of the number of condensed rings *N*. Hexagons symbolize ^1A_g, diamonds $^1B_{2u}$, circles $^1B_{3u}$, and triangles $^1B_{1g}$.

its wave function has very little contributions from single excitations. The 1 ^1A_g → 2 ^1A_g one-photon transition is symmetry-forbidden so that we do not expect to see this state in the absorption spectrum of octacene. The ^1L_a state represents the S_2 state here.

The excitation energy of the ^1L_b state varies significantly less than the one of the ^1L_a and other low-lying singlet states. With extending molecular size, the ^1L_b state therefore switches order quite often (Figure 1). It represents the S_1 state in naphthalene, shifts to S_2 in anthracene and tetracene, and is the third excited singlet (S_3) in pentacene where the doubly excited 2 ^1A_g state takes the position of the S_2 state. As mentioned above, the leading configurations of the 1 $^1B_{3u}$ (^1L_b) state are (HOMO − 2 → LUMO) and (HOMO → LUMO + 2) here. The 1 $^1B_{1g}$ state lies very close by (at 2.99 eV in the SV(P) basis and at 3.02 in the TZVP basis) but is optically dark. Its wave function has large contributions from (HOMO − 1 → LUMO), (HOMO → LUMO + 1), and double excitations. The situation is even more complex in hexacene. As in pentacene, the S_2 state is the doubly excited 2 ^1A_g. The 1 $^1B_{1g}$ has dropped below the ^1L_b state and forms S_3. The 1 $^1B_{3u}$ state is the fourth excited singlet in hexacene. Unlike the situation in the smaller polyacenes, its wave function is dominated by the double excitations (HOMO − 1, HOMO → LUMO^2) and (HOMO^2 → LUMO, LUMO + 1), but has still large components from (HOMO − 2 → LUMO) and (HOMO → LUMO + 2). The latter are the leading configurations in 2 $^1B_{3u}$ which corresponds to S_6 in this molecule. Due to their energetic proximity and mixed wave function character, the two low-lying $^1B_{3u}$ states exhibit similar oscillator strengths and together make up the ^1L_b state. For this reason, two values are listed as ^1L_b excitation energy in Table 1. In octacene, the near-degeneracy of the 1 $^1B_{3u}$ and 2 $^1B_{3u}$ states does not persist. Although the 1 ^1A_g → 1 $^1B_{3u}$ transition at 1.46 eV (SV(P) basis) is formally symmetry-allowed, it is one-photon forbidden because of the double-excitation character of the state. The 2 $^1B_{3u}$ (S_9) wave function clearly represents the ^1L_b state in octacene, closely followed by another optically allowed transition, 1 ^1A_g → 3 $^1B_{2u}$, at 2.86 eV (SV(P) basis).

3.1.2. Short Linear Polyenes. For the first members of the series, gas phase spectra or high-resolution spectra of jet-cooled molecules are available.[39−47] These experimental conditions guarantee that environmental effects on the spectra are small. Electron impact studies[48−51] revealed the approximate positions of triplet states. Furthermore, bond lengths of *trans*-1,3,5-hexatriene have been determined experimentally by gas phase electron diffraction.[52] In addition, low-temperature spectra of short polyenes have been recorded with high resolution in Shpolskii matrices of *n*-alkanes.[53−55] Comparison of the band positions with gas phase spectra can thus provide an estimate of solvent shifts. In addition to numerous experimental investigations, extensive theoretical studies with ab initio wave function methods,[12−14,16,17,56−62] and (TD)DFT[2−4,15,63] have been carried out.

For these reasons, we have chosen *trans*-1,3,5-hexatriene (HT), *trans*,*trans*-1,3,5,7-octatetraene (OT), and *all-trans*-1,3,5,7,9-decapentaene (DP) as benchmark systems for which detailed comparison with experimental data is made. Moreover, these molecules are small enough to test basis set effects on the geometrical parameters and on the spectral properties.

Hexatriene. Conflicting results exist on the energetic position of the 2 ^1A_g state of HT. Buma et al.[44] and Petek et al.[45] have shown by means of fluorescence excitation spectroscopy that a one-photon forbidden state with origin at 4.26 eV (presumably 2 $^1A_1$) lies below the optically allowed 1 $^1B_2$ state in isolated *cis*-hexatriene. The origin of the corresponding 2 ^1A_g of *trans*-hexatriene has not been observed, but it can be assumed to lie within a few hundred wavenumbers of the *cis*-band. Coherent anti-Stokes Raman scattering (CARS) and coherent Stokes Raman scattering (CSRS) experiments on liquid *cis*- and *trans*-hexatriene were carried out at ambient temperatures by Fujii et al.[64] These authors report the two-photon absorption (TPA) band to lie a few thousand inverse centimeters above the intense dipole-allowed one-photon transitions (1 $^1B_2$ ← 1 $^1A_1$ (cis) or 1 ^1B_u ← 1 ^1A_g (trans)). In light of the results of Buma et al.[44] and Petek et al.,[45] the assignment of this TPA band to the 2 $^1A_1$ ← 1 $^1A_1$ or 1 ^1B_u ← 1 ^1A_g appears questionable at first sight. As we shall see below, our calculations can help to resolve this conflict. The position of the 1 ^1B_u ← 1 ^1A_g origin is unambiguous, on the other hand. The origin peak—which is also the strongest peak in the spectrum—is reported at 4.93 eV in gas phase absorption spectra[42,65] and at 4.95 eV in a resonant enhanced multiphoton ionization (REMPI) spectrum.[46] The strong origin transition of the 1 ^1B_u ← 1 ^1A_g band and the lack of extended vibrational structure led Leopold et al.[42] to the conclusion that the geometry of the 1 ^1B_u state is not largely distorted, i.e., in particular it is believed to be planar. Myers and Pranata[66] come to a similar conclusion with regard to planarity but report significant reductions in the force constants for both terminal and central double-bond twisting with respect to the ground state.

Table 2. Basis Set Dependence of Vertical DFT/MRCI Absorption Energies ΔE_{vert} [eV] of HT in Comparison with Previous Theoretical Results of Correlated ab initio Wave Function Methods and Experimental Values

method	DFT/MRCI			QCI/CI6[a]	CASPT2[b,c,d]			CC2[d]	CCSD[d]	CC3[d]	ADC(2)-s[e]	ADC(2)-x[e]	
basis	SV(P)	TZVP	TZVPP										$\lambda_{\max}$
2 ^1A_g	4.95	4.96	4.98	5.74	5.19	5.34	5.42	6.67	6.61	5.72	6.75	4.07	5.21[f]
1 ^1B_u	5.07	4.97	4.94	5.14	5.01	5.37	5.31	5.41	5.72	5.58	5.36	5.15	5.13[g], 4.93[h,i]
1 ^3B_u	2.46	2.47	2.49	2.84	2.55	2.60	2.71	2.78	2.62	2.69			2.61[g]
1 ^3A_g	4.01	3.99	4.01		4.12	4.24	4.31	4.40	4.28	4.32			4.11[g]

[a] See ref 56. QCI for 2 ^1A_g, CI6 + SC for 1 ^1B_u and 1 ^3B_u. [b] See ref 12. CASPT2 based on CASSCF with 6 active electrons in 4 a_u and 4 b_g active orbitals, (6s3p1d/2s1p) ANO basis + Rydberg functions, and experimental geometry parameters. [c] See ref 14. CASPT2 based on CASSCF with 6 active electrons in 12 active orbitals, (3s2p1d/2s) basis, and experimental geometry parameters. [d] See ref 17. CASPT2 based on CASSCF with 6 active electrons in 6 active orbitals, TZVP basis, and ground-state MP2/6-31G* geometry. [e] See ref 6. TZVP basis and ground-state MP2/6-31G* geometry. [f] See ref 64. TPA maxiumum of liquid HT at room temperature. [g] See ref 49. Band maximum of low-energy electron impact spectrum. [h] See ref 42. 0−0 transition and absorption maximum of jet-cooled HT. [i] See ref 39. 0−0 transition and absorption maximum of isolated HT.

Electron impact studies[48,49] observe a different intensity distribution of the transition. Flicker et al.[49] find the 0−0 transition at 4.95 eV in agreement with optical spectra but the band maximum occurs at 5.13 eV. Despite considerable effort no emission has been observed from *trans*-1,3,5-hexatriene after 1 $^1B_u \leftarrow 1\ ^1A_g$ excitation. According to Leopold et al.,[42] the observed broadband widths of the hexatriene absorption spectrum appear to be compatible with an extremely short lifetime of the primarily excited state due to ultrafast internal conversion to a lower-lying singlet. Transitions to higher-lying singlet states of HT were observed by Gavin and Rice[39] but were assigned to Rydberg transitions. Since we did not include diffuse basis functions, a comparison with experiment cannot be made in these cases.

Due to different selection rules, transitions to triplet states can be observed with high intensity by means of electron impact spectroscopy. Vertical excitation energies were published for T_1 (2.61 eV, 1 ^3B_u) and T_2 (4.11 eV, 1 ^3A_g).[49] Frueholz and Kuppermann[50] could resolve the vibronic structure of the second triplet and were thus able to determine its adiabatic excitation energy (3.75 eV).

The vertical DFT/MRCI excitation energies of HT (Table 2) show a similar trend as already observed for the polyacenes. For the two lowest singlet states, the experimental reference values are underestimated by about 0.2 eV, but no preference is given to either the 2 ^1A_g or the 1 ^1B_u state. Both are found to be nearly degenerate in the Franck−Condon (FC) region, in agreement with CASPT2 results[12,14,17] and experimental evidence.[42] In contrast, the coupled-cluster methods with single and double excitation operators (CCSD and CC2)[17] and the strict ADC(2) approach, ADC(2)-s,[6] yield very large energy gaps. It appears that they do not properly account for the multiconfiguration effects in the 2 ^1A_g state. In HT, the (HOMO − 1 → LUMO) and (HOMO → LUMO + 1) single excitations are prominent configurations in the CI expansion of the S_1 state, but unlike the situation in short polyacenes, they are not the leading terms. The latter is dominated by the (HOMO → LUMO)^2 double excitation instead. The extended ADC(2) method, ADC(2)-x, which includes the treatment of double excitations through first order, seems to overshoot, yielding a significantly too low excitation energy of the 2 ^1A_g state in HT.[6] With regard to the basis set dependence of the DFT/MRCI results, only slight variations of the 2 ^1A_g excitation energy are observed. The 1 ^1B_u state, which originates from the (HOMO → LUMO) excitation and corresponds to ionic valence-bond structures, appears to be more sensitive to the quality of the basis set. The two lowest triplet states are mainly represented by single excitations. As observed already by Schreiber et al.[17] their correlation treatment is less demanding. Here, all the methods perform equally well.

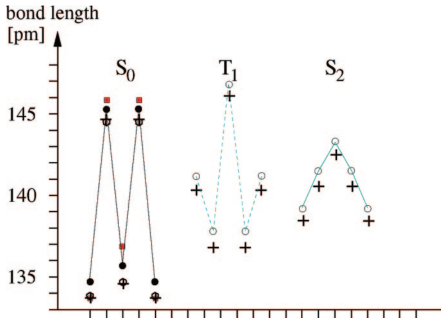


Figure 2. C−C bond lengths of *trans*-1,3,5-hexatriene in the 1 ^1A_g electronic ground state (left), the first excited triplet state 1 ^3B_u (middle), and the optically bright 1 ^1B_u state (right). Calculated values are represented by circles (B3-LYP, SV(P) basis), plus signs (B3-LYP, TZVP basis), and hexagons (B3-LYP, TZVPP basis). Squares correspond to experimentally derived values determined from gas phase electron diffraction.[52]

In order to distinguish direct basis set effects on the excitation energy and indirect effects through the geometry, we carried out single-point DFT/MRCI calculations in the SV(P) basis set at the ground-state geometries obtained from B3-LYP optimizations in the TZVP and TZVPP bases, respectively, for the wave function methods HF, RIMP2, and MP2 and for a structure with experimentally derived bond distances.[52] The TZVP- and TZVPP-optimized geometries are practically identical while the bond distances of the SV(P)-optimized structure are consistently longer by about 1 pm (see Figure 2). Comparison to the geometry parameters derived from the X-ray structure of gaseous HT[52] shows that the terminal double bond length is reproduced excellently in the calculations using at least a TZVP basis. Our value for the central double bond distance (134.7 pm) is in good agreement with the CASSCF value of 134.5 pm by Nakayama et al.[14] but is markedly shorter than the experimental value (136.7 pm).[52] We refrained from optimizing the geometry of the 2 ^1A_g state because of its double excitation character. For the T_1 state (1 ^3B_u), a reversal of single and

Table 3. Calculated Adiabatic DFT/MRCI Excitation Energies ΔE_{adia} [eV] of HT in Comparison with Experimental Values

	ΔE_{adia} (DFT/MRCI)		
state	SV(P)	TZVP	ΔE_{0-0}
2 ^1A_g	4.07[a]	4.08[a]	4.26[b]
1 ^1B_u	4.79	4.67	4.93[c, d], 4.94[e], 4.95[f]
1 ^3B_u	1.90	1.94	

[a] Energy at the 1 ^3B_u minimum. [b] See ref 45. Fluorescence excitation of jet-cooled *cis*-1,3,5-hexatriene. [c] See ref 65. Absorption of isolated HT. [d] See ref 42. Absorption of jet-cooled HT. [e] See ref 46. REMPI spectrum. [f] See ref 49. Low-energy electron impact.

double bond character is found with the difference between double and single bond being less pronounced than in the ground state. The bond length alternation is even smaller in the 1 ^1B_u state. In both cases, the trends for the SV(P) and TZVP bases are the same as for the ground state, i.e., the SV(P) basis yields minimum nuclear structures with slightly longer bonds. Detailed results of the single-point calculations are available in the Supporting Information (SI). In all cases, the DFT/MRCI excitation energy obtained with the SV(P) basis at the SV(P)-optimized geometry agrees much better with the TZVP and TZVPP results than the DFT/MRCI value computed in the SV(P) basis at the TZVP- and TZVPP-optimized nuclear structures. We therefore conclude that it is advantageous to employ the same basis set for geometries and excitation energies.

Proceeding finally to adiabatic excitation energies (Table 3), few data are available for comparison. For the reasons discussed above, an optimized minimum geometry of the 2 ^1A_g state is not easily obtained. In the absence of the appropriate nuclear arrangement, we noticed that the absolute energy of the 2 ^1A_g takes the lowest value at the T_1 (1 ^3B_u) minimum geometry which is conceivable on the basis of qualitative arguments along the line of Walsh rules. For the 1 ^3B_u and 1 ^1B_u states, the UDFT- and TDDFT-optimized nuclear arrangements were employed, respectively. We find that the geometry relaxation effect on the excitation energy is much more pronounced for the 2 ^1A_g state than for the 1 ^1B_u state. Even at the relaxed 1 ^1B_u minimum geometry, we find the 2 ^1A_g state to lie more than 0.5 eV below the 1 ^1B_u state. This explains the seemingly conflicting obvervations by Buma et al.[44] and Petek et al.[45] on the one side and by Fujii et al.[64] on the other side and makes broad band widths of the hexatriene absorption spectrum[42] plausible. It appears that the two states are near degenerate in the FC region with the 2 ^1A_g state possibly located slightly above the optically bright 1 ^1B_u state. Upon geometry relaxation, the 2 ^1A_g state clearly becomes the S_1 state, opening a fast relaxation channel for the primarily occupied 1 ^1B_u state. Following the general trends, the DFT/MRCI calculations underestimate the adiabatic excitation energies of both states by about 0.2−0.3 eV. For the T_1 state we did not find a published value of the 0−0 energy, but we expect the deviation to be slightly smaller than in the singlet cases.

Octatetraene. In order to assess the results of the present theoretical study on OT, a short review of the experimental findings will be given first. Gas phase OT has been reported to fluoresce from its S_2 (1 ^1B_u) state with a quantum yield of about 0.1.[40] The origin of the 1 ^1B_u ← 1 ^1A_g transition of jet-cooled OT is observed at 4.41 eV.[41,43] Since it represents the strongest peak in the absorption spectrum of OT, this value is frequently used also by theoreticians as a measure for the vertical excitation energy at the ground-state equilibrium geometry.[12,14,17] For lack of an alternative, we will do the same here, but we regard the comparison with some reservation. As detailed by Davidson and Jarzecki,[67] a comparison of the computed vertical excitation energy with the corresponding maximum in the absorption spectrum implicitly makes the assumption that the geometry displacement of the excited state is fairly large so that its vibrational wave function has maximal amplitude in the region of the classical turning points, which is not the case here. Gavin et al.[40] estimate the energy gap between the 1 ^1B_u and 2 ^1A_g state of OT to be approximately 0.8 eV, in line with the results of newer measurements. In 1999, Pfanstiel et al.[47] succeeded in recording a rotationally resolved one-photon fluorescence excitation spectrum of OT. The first strong band at 3.60 eV is located only marginally above the origin of the 2 ^1A_g ← 1 ^1A_g emission at 3.59 eV. In the condensed phase at 77 K, relaxed emission from the lower-lying 2 ^1A_g state of OT is observed with a quantum yield of approximately 0.6.[40] The origin transition of 2 ^1A_g ← 1 ^1A_g was determined by TPA in two different solvents.[54] The solvent shift of this transition is rather small. In contrast, the 1 ^1B_u state experiences large stabilization effects by the surrounding *n*-alkanes, decreasing the energy gap between the 1 ^1B_u and 2 ^1A_g state in *n*-hexane to approximately half of the gas phase value. This should be kept in mind when comparing quantum chemical results for isolated molecules with experimental data in the condensed phase.

Table 4. C−C Bond Distances [pm] in the Electronic Ground State of OT

	B-LYP[a]	B3-LYP[a]	BH-LYP[a]	HF[a]	MP2[a]	CASSCF[b]	exp[c]
C_4−C_3	136.1	134.8	133.4	132.9	135.1	134.5	133.6
C_3−C_2	145.3	145.0	144.9	146.4	145.1	145.7	145.1
C_2−C_1	137.5	136.0	134.3	133.5	136.1	135.1	132.7
C_1−$C_{1'}$	144.4	144.3	144.4	146.0	144.5	145.1	145.1

[a] Present work. SV(P) basis. [b] See ref 14. CASSCF with 8 active electrons in 12 active orbitals, (3s2p1d/2s) basis. [c] See ref 68. X-ray structure of crystalline OT.

Detailed information on the vertical and adiabatic excitation energies of the two lowest triplet states is available through electron energy loss spectra which were recorded with resolution of the vibrational structure by Allan et al.[51] The authors also observed the 1 ^1B_u ← 1 ^1A_g transition and a higher-lying band with origin and maximum at 6.04 eV which they tentatively assigned to the 2 ^1B_u ← 1 ^1A_g transition, but state that it could be due to the 3 ^1A_g state.

The results of the present theoretical investigation on OT are collected in Tables 4−6. The minimum nuclear arrangement of the electronic ground state was optimized using three different density functionals as well as HF and MP2. The resulting C−C bond lengths are displayed in Table 4 together with earlier theoretical values obtained at the CASSCF level[14] and crystal structure data.[68] It must be stressed again that in all cases the MO basis and the Fock matrix elements

Table 5. Dependence of DFT/MRCI Vertical Absorption ΔE_{vert} and Adiabatic Excitation Energies ΔE_{adia} [eV] of OT on the Geometry Parameters

	ΔE_{vert}, SV(P) basis							ΔE_{adia}, SV(P) basis			
geometry	B-LYP	B3-LYP	BH-LYP	HF	MP2	exp[a]	λ_{max}	B-LYP	B3-LYP	BH-LYP	ΔE_{0-0}
2 ^1A_g	3.82	4.02	4.26	4.48	4.01	4.41	≈4.1[b]	3.21[c]	3.25[c]	3.29[c]	3.59[d], 3.54[e], 3.56[f]
1 ^1B_u	4.24	4.33	4.45	4.57	4.33	4.54	4.40[g]	4.12	4.10	4.07	4.41[h], 4.41[i], 4.40[g]
2 ^1B_u	5.08	5.28	5.53	5.69	5.27	5.67					
3 ^1A_g	5.81	6.00	6.22	6.32	5.98	6.30					
1 ^3B_u	1.90	2.02	2.17	2.33	2.02	2.30	2.10[g]	1.53	1.53	1.51	1.73[g]
1 ^3A_g	3.21	3.32	3.44	3.50	3.30	3.46	3.55[g]				3.25[g]
2 ^3B_u	4.31	4.42	4.54	4.57	4.41	4.57					
2 ^3A_g	5.04	5.15	5.26	5.27	5.13	5.34					
3 ^3B_u	5.30	5.40	5.80	5.87	5.48	5.80					

[a] C−C bond distances adjusted to experimental values taken from ref 68. Crystal structure. [b] See ref 46. Estimated as 0−0 energy + 0.5 eV. [c] Energy at the 1 ^3B_u minimum. [d] See ref 47. Fluorescence excitation of jet-cooled OT. [e] See ref 54. Two-photon absorption in *n*-octane. [f] See ref 54. Two-photon absorption in *n*-hexane. [g] See ref 51. Electron energy loss of gaseous HT. [h] See ref 41. Absorption of jet-cooled OT. [i] See ref 43. Absorption and emission of jet-cooled OT.

Table 6. Basis Set Dependence of Calculated Vertical Absorption Energies ΔE_{vert} [eV] of OT and Comparison with Previous Theoretical Results of Correlated ab initio Wave Function Methods and Experimental Values

	DFT/MRCI			QCI/CI8[a]	CASPT2[b,c,d]			CC2[d]	CCSD[d]	CC3[d]	
state	SV(P)	TZVP	TZVPP								λ_{max}
2 ^1A_g	4.02	4.06	4.08	5.21	4.38	4.72	4.64	5.87	5.99	4.97	≈4.1[e]
1 ^1B_u	4.33	4.27	4.24	4.79	4.42	4.81	4.70	4.71	5.07	4.94	4.41[f], 4.41[g], 4.40[h]
2 ^1B_u	5.28	5.27	5.29			5.76	5.74	6.91	6.89	6.06	
3 ^1A_g	6.00	5.82	5.85			6.40	6.19	6.72	6.98	6.50	
1 ^3B_u	2.02	2.05	2.06	2.52	2.17	2.37	2.33	2.40	2.23	2.30	2.10[h]
1 ^3A_g	3.32	3.31	3.36		3.39	3.61	3.70	3.76	3.62	3.67	3.55[h]
2 ^3B_u	4.42	4.39	4.42			4.71					
2 ^3A_g	5.15	5.02	5.05			5.43					
3 ^3B_u	5.40	5.48	5.51								

[a] See ref 58. QCI for 2 ^1A_g, CI8 + SC for 1 ^1B_u and 1 ^3B_u. [b] See ref 13. CASPT2 based on CASSCF with 8 active electrons in 6 a_u and 5 b_g active orbitals, (4s3p1d/2s1p) ANO basis + Rydberg functions, experimental geometry parameters. [c] See ref 14. CASPT2 based on CASSCF with 8 active electrons in 12 active orbitals, (3s2p1d/2s) basis, experimental geometry parameters. [d] See ref 17. CASPT2 based on CASSCF with 8 active electrons in 8 active orbitals, TZVP basis, ground-state MP2/6-31G* geometry. [e] See ref 46. Estimated as 0−0 energy + 0.5 eV. [f] See ref 41 absorption of jet-cooled OT. [g] See ref 43. Absorption and emission of jet-cooled OT. [h] See ref 51. Electron energy loss of gaseous HT.

of the DFT/MRCI-Hamiltonian are constructed from a Kohn−Sham calculation employing the BH-LYP functional.

The performance of the B-LYP functional is interesting under cost considerations since the possibility to employ the RI approximation brings about an enormous speed-up of the geometry optimization compared to the hybrid functionals. The C−C bond length alternation of the B-LYP ground-state structure (no HF exchange) is less pronounced than in the B3-LYP minimum geometry (20% HF exchange) whereas the contrary is true for the BH-LYP functional (50% HF exchange). Comparing with the pure HF results it appears that a higher percentage of HF exchange leads primarily to a shortening of the double bonds while the single bond length variation is less pronounced. Addition of electron correlation, on the other hand, causes a decrease of the bond length alternation. The MP2- and B3-LYP-optimized geometries are nearly indistinguishable. These C−C bond lengths also closely resemble the CASSCF values published by Nakayama et al.[14] The X-ray values for the C−C single bond lengths[68] are comparable to the theoretical values. However, as already noticed by Serrano-Andrés et al.[12] and Nakayama et al.,[14] the X-ray value for the double bond adjacent to the central single bond in crystalline OT (132.7 pm)[68] is presumably too short.

For the study of the geometry dependence of vertical and adiabatic excitation energies (Table 5), we employed the SV(P) basis throughout. A less pronounced bond length alternation introduces a bias in favor of the excited states. Accordingly, the vertical DFT/MRCI excitation energies, computed at the B-LYP geometry, are consistently lower than at the B3-LYP geometry by about 0.2 eV for doubly excited states and about 0.1 eV for single excitations. Again the reverse trend is observed at the BH-LYP geometry and is continued for the HF geometry. The vertical DFT/MRCI excitation energies at the BH-LYP geometry are in good agreement with the experimental data, with the above-mentioned reservation that the position of the origin of the 1 $^1B_u \leftarrow 1\ ^1A_g$ transition is not a good measure for the vertical excitation energy even if this is the strongest band in the vibrationally resolved spectrum. It might be worth noting in this context that in the parametrization of the DFT/MRCI-Hamiltonian against experimental band maxima originally BH-LYP optimized ground-state geometries were employed.[21] The general trend of the DFT/MRCI method to underestimate the true excitation energy might therefore partially result from a geometry effect during the parametrization. The decision, not to use the BH-LYP functional for geometry optimization is made on the basis of two facts: (1) The B3-LYP optimized geometry yields the lowest

Table 7. Calculated Vertical Absorption Energies ΔE_{vert} and Adiabatic Excitation Energies ΔE_{adia} [eV] of DP in Comparison with Previous Theoretical Results of Correlated ab initio Wave Function Methods and Experimental Values

	ΔE_{vert} (DFT/MRCI)			CASPT2[a]	ΔE_{adia} (DFT/MRCI)		
state	SV(P)	TZVP	TZVPP		SV(P)	TZVP	ΔE_{0-0}
2 ^1A_g	3.40	3.44	3.46	3.95	2.68[b]	2.71[b]	3.07[c], 3.05[d], 3.10[e]
1 ^1B_u	3.82	3.77	3.74	3.97	3.61	3.54	3.57[f,g], 3.98[e], 4.02[h]
2 ^1B_u	4.55	4.57	4.59	4.91			
3 ^1A_g	5.38	5.32	5.31	5.64			
1 ^3B_u	1.73	1.76	1.77	1.95	1.28	1.37	
1 ^3A_g	2.82	2.82	2.84	3.02			
2 ^3B_u	3.85	3.83	3.86	4.07			
2 ^3A_g	4.65	4.59	4.63	4.86			
3 ^3B_u	4.69	4.70	4.73	4.97			

[a] See ref 14. CASPT2 based on CASSCF with 10 active electrons in 10 active orbitals, (3s2p1d/2s) basis, and experimental geometry parameters. [b] Energy at the 1 ^3B_u minimum. [c] See ref 92. Two-photon absorption in *n*-heptane at 77 K. [d] See ref 92. Two-photon absorption in *n*-decane at 77 K. [e] See ref 69. 0−0 band measured in various solvents, corrected for solvent shifts. [f] See ref 92. Absorption and emission in *n*-heptane at 77 K. [g] See ref 92. Absorption and emission in *n*-decane at 77 K. [h] See ref 69. Origin of gas phase absorption.

absolute DFT/MRCI energy of the electronic ground state. (2) TDDFT calculations on the lowest triplet of longer polyenes are prone to triplet instabilities when the BH-LYP functional is employed. Interestingly, the influence of the particular choice of density functional for the geometry optimization levels off in case of the adiabatic DFT/MRCI excitation energies. Obviously, a partial cancelation of geometry effects occurs here.

Fortunately, the sensitivity of the DFT/MRCI energies with respect to the basis set (Table 6) is much less pronounced than the geometry dependence discussed above. As already seen for the polyacenes, an SV(P) basis appears to be sufficient. The DFT/MRCI method yields the correct order of low-lying singlet states. Admittedly, the 4.1 eV referred to as vertical excitation energy of the 2 ^1A_g state by McDiarmid[46] is only a rough estimate. On the other hand, we know that the 0−0 energy of the 1 $^1B_u \leftarrow 1\ ^1A_g$ transition is definitely a lower limit of the true vertical excitation energy. It is therefore clear that the 2 ^1A_g state is located below the 1 ^1B_u state in the FC region or that they are at most near degenerate. From the ab initio methods only the CASPT2 finds the 2 ^1A_g state as the S_1 state. The CC3 method, which was employed by Schreiber et al.[17] for benchmark reasons only and is too expensive to be used in practical applications on large polyenes, obtains near degeneracy of the 1 ^1B_u and 2 ^1A_g states whereas the less demanding CC2 approach places the 2 ^1A_g more than 1.1 eV above the 1 ^1B_u state. Similarly large energy gaps with the wrong ordering of states are found for (TD)DFT treatments using hybrid functionals.[4]

Decapentaene. For *all-trans*-1,3,5,7,9-decapentaene, less experimental data are available. Vibrationally resolved fluorescence and one-photon excitation as well as TPA have been recorded in glassy *n*-alkanes.[55] A reliable estimate of the S_1 and S_2 origins was presented by D'Amico et al.[69] The formula which they employed for the extrapolation, $\nu(\text{solvent}) = \nu(\text{gas}) - k(n^2 - 1)/(n^2 + 2)$ where k is a fitting parameter and n is the refractive index of the solvent, yielded results in good agreement with gas phase data in the case of OT and the S_2 state of DP.

The DFT/MRCI results obtained for different AO bases (Table 7) are consistent with the findings in HT and OT. The vertical excitation energies are systematically lower than the CASPT2 results by Nakayama et al.,[14] but trends are well reproduced. A comparison with experiment can be made only for the adiabatic transitions between the ground state and the two lowest singlet excited states. The estimated S_1 and S_2 energies of isolated DP are underestimated by about 0.4 eV in this case, but the energy gap is in the right ballpark.

3.1.3. α, ω-Diphenyl-Polyenes. Similar trends are found for the short α, ω-diphenyl-polyenes, 1,6-diphenyl-*trans*, *trans*-1,3,5-hexatriene (DPHT), 1,8-diphenyl-*all-trans*-1,3,5,7-octatetraene (DPOT), 1,10-diphenyl-*all-trans*-1,3,5,7,9-decapentaene (DPDP), and 1,12-diphenyl-*all-trans*-1,3,5,7,9,11-dodecahexaene (DPDH). At first sight (Table 8), one might think that the deviations from the experimental excitation energies of the 1 $^1B_u \leftarrow 1\ ^1A_g$ transition are smaller. However, for these molecules, only spectra in the condensed phase have been measured[7,70−74] and solvent effects are known to preferentially stabilize the 1 ^1B_u state. In CS_2, a highly polarizable agent, even an inversion of the 1 ^1B_u and 2 ^1A_g levels has been observed.[73]

With regard to the cost of calculations on carotenoids, we investigated the basis set dependence and the sensitivity with respect to the density functional used for the geometry optimization. The same trends as for the short linear polyenes are found: The changes are very small when proceeding from the TZVPP basis over the TZVP basis to the SV(P) basis, whereas red shifts of the order of 0.1−0.2 eV are observed when the B-LYP geometry is employed instead of the B3-LYP geometry of the ground state. Detailed results are provided in Table 2 of the SI.

3.2. Longer Polyenes. From our benchmark results on shorter polyenes, α, ω-diphenyl-polyenes, and polyacenes and from recent results of the Thiel group,[17] we expect the DFT/MRCI energies at the B3-LYP optimized geometries to systematically underestimate the true energies of the electronically excited states. We did, however, not observe a bias toward one or the other type of state. It can therefore be expected that the energy gaps between the electronically excited states are reproduced well. Information on the performance of the DFT/MRCI method with respect to CPU time and on the number of CSFs included in the final MRCI space is available in the SI.

Table 8. DFT/MRCI Vertical Excitation Energies ΔE_{vert} and Adiabatic Excitation Energies ΔE_{adia} [eV] of α, ω-Diphenyl-polyenes in Comparison with Experimental Band Origins[a]

	ΔE_{vert} (DFT/MRCI)		ΔE_{adia} (DFT/MRCI)		
molecules	SV(P)	TZVP	SV(P)	TZVP	ΔE_{0-0}
			2 ^1A_g state[b]		
DPHT	3.47	3.51	2.86	2.87	3.19[c], 3.12[d]
DPOT	3.00	3.05	2.42	2.44	2.80[e], 2.77[d,f]
DPDP	2.64	2.70	2.08	2.11	2.50[g]
DPDH	2.35	2.42	1.84	1.84	2.26[g]
			1 ^1B_u state		
DPHT	3.36 (2.21)	3.34 (2.18)	3.15	3.10	3.62[c], 3.22[d], 3.23[d]
DPOT	3.08 (2.66)	3.07 (2.65)	2.88	2.85	3.01[e], 2.96[d], 3.02[f]
DPDP	2.85 (3.08)	2.85 (3.09)	2.66	2.64	2.86[g]
DPDH	2.66 (3.48)	2.67 (3.50)	2.47	2.46	2.72[g]
			1 ^3B_u state		
DPHT	1.78	1.81	1.38	1.39	1.50[h]
DPOT	1.56	1.59	1.17	1.19	
DPDP	1.39	1.43	1.02	1.04	
DPDH	1.25	1.29	0.89	0.91	

[a] Oscillator strengths for dipole-allowed transitions from the ground state are displayed in parentheses. [b] DFT/MRCI excitation energy at the 1 ^3B_u minimum. [c] See ref 73. Extrapolated vacuum origin. [d] See ref 71. Two-photon absorption at 77 K in EPA. [e] See ref 7. Emission spectrum recorded at 4.2 K in pentadecane. [f] See ref 70. Two-photon absorption at 77 K in EPA. [g] See ref 72. Excitation (1 ^1A_g → 1 ^1B_u) and fluorescence (2 ^1A_g → 2 ^1A_g) spectra recorded at 4.2 K in *n*-decane (DPDP) or *n*-dodecane (DPDH). [h] See ref 74. Triplet excitation spectra of DPHT crystals at 20 K.

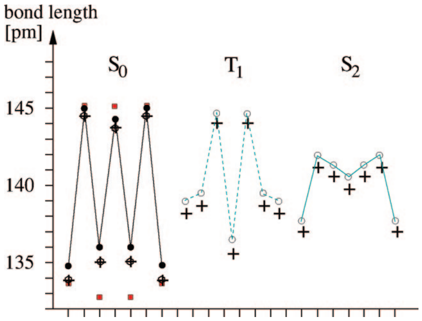


Figure 3. C−C bond lengths of *trans,trans*-1,3,5,7-octatetraene in the 1 ^1A_g electronic ground state (left), the first excited triplet state 1 ^3B_u (middle), and the optically bright 1 ^1B_u state (right). Calculated values are represented by circles (B3-LYP, SV(P) basis), plus signs (B3-LYP, TZVP basis), and hexagons (B3-LYP, TZVPP basis). Squares correspond to experimentally derived values determined from X-ray studies on crystalline OT.[68]

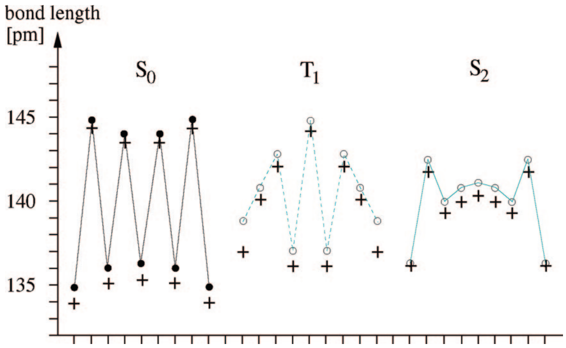


Figure 4. Calculated C−C bond lengths (circles: B3-LYP, SV(P) basis; plus signs: B3-LYP, TZVP basis) of *all-trans*-1,3,5,7,9-decapentaene in the 1 ^1A_g electronic ground state (left), the first excited triplet state 1 ^3B_u (middle), and the optically bright 1 ^1B_u state (right).

3.2.1. Vertical Excitation at the Ground State Geometry. The ground-state equilibrium structure is characterized by alternating double and single bonds wherein the bond length alternation is largest at the polyene ends and decreases slightly toward the center (Figure 5, top). The 1 ^3B_u state, which is dominated by the (HOMO → LUMO) single excitation, constitutes the lowest excited state in all polyenes. In the simple model of $2N$ independent electrons in a one-dimensional box of width L the excitation energy is given by the HOMO−LUMO orbital energy gap which is proportional to $(2N + 1)/L^2$. It is customary to assume that the box potential extends further than the distance between the first and last carbon atom which is approximately equal to $(2N - 1)$ times the average C−C bond length R_{CC}. Adding one bond length on either side of the polyene chain yields an estimate for the box width $L \propto (2N + 1)R_{\mathrm{CC}}$. Within this model, a straight line would result if the vertical excitation energy is plotted as function of $1/(2N + 1)$. It is seen (Figure 6) that the excitation energies of the T_1 state nicely follow this simple scheme. The T_2 state, which comes next in the vertical excitation spectrum, possesses ^3A_g symmetry and has two leading configurations, (HOMO − 1 → LUMO) and (HOMO → LUMO + 1). With regard to its conjugation length dependence, a steeper slope and a larger deviation from linearity is observed in comparison to the T_1 state. This effect is even more pronounced for the corresponding singlet state, 2 ^1A_g. In addition to the above-mentioned (HOMO − 1 → LUMO) and (HOMO → LUMO + 1) single excitations, double excitations contribute to the CI expansion with high weight wherein the (HOMO → LUMO)^2 configuration is the leading term, in accord with earlier semiempirical calculations.[9,10] For the longer polyenes, also significant admixture with the double excitation (HOMO − 1, HOMO → LUMO, LUMO + 1) and the ground-state configuration is found. According to our calculations, this multiconfigurational expansion yields the lowest excited singlet state in all linear polyenes with $N \geq 3$, in agreement with experimental evidence.[7,8] Calculated vertical excitation energies of the S_1 state and the optically bright ^1B_u state, which results mainly from the (HOMO → LUMO) single excitation, can be found in Table 9. For all but the longest polyenes, the latter state is the second excited state at the ground-state

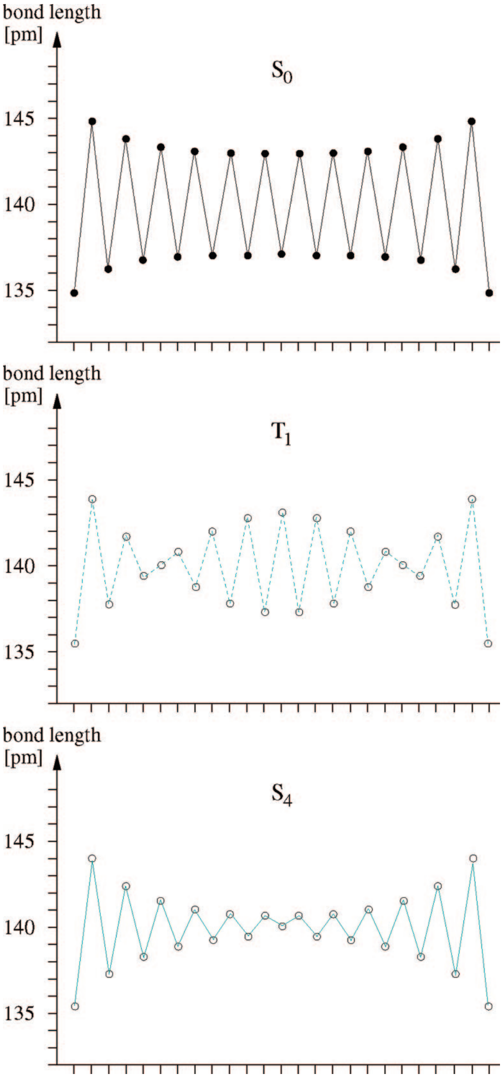


Figure 5. Calculated C−C bond lengths (B3-LYP, SV(P) basis) of *all-trans*-1,3,5,7,9,11,13,15,17,19,21,23,25-hexacosatridecaene in the 1 ^1A_g electronic ground state (top), the first excited triplet state 1 ^3B_u (middle), and the optically bright 1 ^1B_u state (bottom).

geometry. As discussed before, 1 ^1B_u is nearly degenerate with the 2 ^1A_g state and the 2 ^3B_u state in the vertical excitation spectrum of HT. The second singlet and triplet B_u states, symbolized by circles in Figure 6, and the third triplet B_u, symbolized by upside down triangles, have similarly strong multiconfigurational character as the 2 ^1A_g state. In the shorter polyenes the (HOMO − 2 → LUMO), (HOMO → LUMO + 2), and (HOMO − 1 → LUMO + 1) dominate these states while in the longer polyenes the doubly excited (HOMO − 1, HOMO → LUMO^2) and (HOMO^2 → LUMO, LUMO + 1) gain weight and become predominant in the 2 ^1B_u and 3 ^3B_u states. It is noteworthy that the order of the ^1B_u states changes with conjugation length. The optically bright ^1B_u state, frequently labeled $^1B_u^+$ in the literature wherein the + sign represents the so-called Pariser alternancy symmetry,[75] is the second excited state in the vertical excitation spectrum of short polyenes. According to our calculations, it becomes nearly degenerate with the second ^1B_u state, also denominated $^1B_u^-$ in the literature, for N = 11 while $^1B_u^+$ represents the third excited singlet state for longer polyenes. As we will see below, the location of the crossover is geometry dependent though. Let us finally turn to the 2 ^3A_g and 3 ^1A_g states which are the last ones that we have analyzed in detail. At the ground-state geometry, the 2 ^3A_g state is mainly composed of the single excitations (HOMO − 3 → LUMO), (HOMO − 2 → LUMO + 1), (HOMO → LUMO + 3), and (HOMO − 1 → LUMO + 2). The corresponding singlet, 3 ^1A_g, has significant contributions from the double excitations (HOMO − 2, HOMO → LUMO^2) and (HOMO^2 → LUMO, LUMO + 2) which become the leading terms in the longer polyenes.

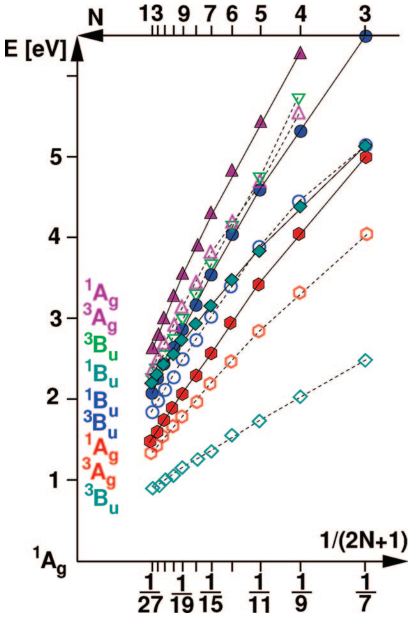


Figure 6. DFT/MRCI vertical electronic excitation energies of *all-trans*-polyenes at the respective 1 ^1A_g ground-state minimum geometry as functions of the conjugation length *N*. Filled symbols and solid lines correspond to singlet states, and open symbols and dashed lines, to triplet states. Squares, hexagons, and upside triangles symbolize A_g symmetric states; diamonds, circles, and upside down triangles symbolize B_u states.

3.2.2. Geometry Relaxation in the T_1 State. In the T_1 state, a reversal of single and double bond character is observed in the central parts of the polyenes (see Figures 2−4 for HT, OT, and DP, respectively, Figure 5 for *all-trans*-1,3,5,7,9,11,13,15,17,19,21,23,25-hexacosatridecaene (HCTD), and the SI for all other polyenes). The differences between alternating bond lengths become smaller when proceeding outward until subsequent bonds are nearly equal. This inner region comprises N − 2 bonds for odd N and N − 1 bonds for even N. Execpt for HT, which is too short to comply with this pattern, the inner region is followed by two (in the polyenes with even N) or three (in the polyenes with odd N) nearly equally long bonds. In the outer part, the bond alternation resembles the one in the ground state so that the terminal bond is always a short one, again except for the smallest members HT, OT, and DP.

This pattern was controversially discussed in the literature. Early studies by Kuki at al.[76] based on semiempirical Pariser−Parr−Pople single and double CI (SDCI) calculations observed strong bond alternation at the terminating carbon atoms and a loss of it in the central part of the polyene

Table 9. Vertical Absorption ΔE_{abs} and Emission ΔE_{em} Energies As Well As Adiabatic Excitation Energies ΔE_{adia} [eV] of Linear Polyenes with Conjugation Length N[a]

	$2\ ^1A_g$			$1\ ^1B_u/2\ ^1B_u$[b]			$1\ ^3B_u$		
N	ΔE_{abs}	ΔE_{em}[c]	ΔE_{adia}[c]	ΔE_{abs} ($f(r)$)	ΔE_{em} ($f(r)$)	ΔE_{adia}	ΔE_{abs}	ΔE_{em}	ΔE_{adia}
3	4.95	3.60	4.07	5.07 (1.39)	4.59 (1.33)	4.21	2.46	1.43	1.90
4	4.02	2.85	3.25	4.33 (1.82)	3.96 (1.77)	4.10	2.02	1.13	1.53
5	3.40	2.34	2.68	3.82 (2.24)	3.50 (2.20)	3.61	1.73	0.94	1.28
6	2.92	1.96	2.26	3.43 (2.64)	3.15 (2.60)	3.23	1.52	0.80	1.10
7	2.56	1.68	1.93	3.14 (3.02)	2.89 (2.98)	2.95	1.36	0.70	0.96
8	2.29	1.46	1.68	2.90 (3.40)	2.67 (3.15)	2.72	1.24	0.64	0.86
9	2.06	1.30	1.50	2.70 (3.76)	2.50 (3.72)	2.53	1.14	0.58	0.78
10	1.88	1.18	1.35	2.54 (4.11)	2.35 (4.08)	2.37	1.07	0.54	0.71
11	1.72	1.08	1.22	2.40 (4.30)	2.22 (4.44)	2.23	1.00	0.52	0.66
12	1.60	1.01	1.12	2.29 (4.78)	2.11 (4.79)	2.10	0.95	0.51	0.62
13	1.49	0.94	1.03	2.18 (5.10)	2.01 (5.14)	2.00	0.91	0.50	0.59

[a] Oscillator strengths $f(r)$ of symmetry-allowed vertical transitions are displayed in parentheses. [b] Ground state geometry: $1\ ^1B_u$ for $N \leq 10$, $2\ ^1B_u$ for $N \geq 11$; excited-state geometry $1\ ^1B_u$ for $N \leq 7$, $2\ ^1B_u$ for $N \geq 8$. [c] Relaxed geometry corresponds to the $1\ ^3B_u$ minimum.

chain instead. The authors related this central domain of diminishing bond alternation to bond orders and postulated a "triplet-excited" region. Subsequent investigations by Takahasi et al.[77,78] pointed out the deficiencies in this approach: Their single-excitation CI (SCI) calculations on the series of polyenes from C_8H_{18} to $C_{22}H_{46}$ yielded geometries in agreement with Kuki et al. However, comparison calculations with CASSCF on octatetraene gave a considerably different bond alternation pattern which is consistent with our present calculations. Takahashi et al. concluded the appearance of the postulated "triplet-excited region" to be artifactual in SCI and SDCI calculations, related to the use of RHF MOs in combination with limitations on the excitations in the CI space. Consideration of the higher polyenes employing this method was not feasible, however. Geometry optimizations by Ma et al.[79,80] focusing on an assessment of the semiempirical Pariser−Parr−Pople model as well as providing results obtained with the UB3LYP method observed a bond alternation pattern which was qualitatively consistent with our study.

Inspection of the T_1 geometry pattern poses the question of a possible electron localization in the two regions of diminishing bond length. While Kuki et al. related their (erroneously established) domain of diminishing bond alternation to bond orders, Takahashi et al. as well as Ma et al. refrained from further conclusions in this respect in their later work, confining themselves to an observation of geometrical effects. Current work in our group concerning this particular question is underway.

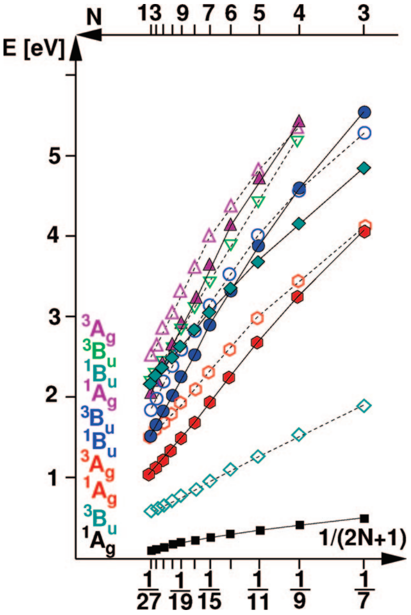


Figure 7. DFT/MRCI vertical electronic excitation energies of *all-trans*-polyenes at the first excited triplet $1\ ^3B_u$ (HOMO → LUMO) minimum geometry as functions of the conjugation length N. The DFT/MRCI energy of the respective $1\ ^1A_g$ ground-state minimum has been chosen as energy offset. For an explanation of symbols, see Figure 6.

The geometry relaxation effect on the excitation energies (Figure 7) is quite dramatic for some of the states as may be expected for a bond-order reversal in the central part of the chromophore. For the short polyenes, the energy gain in the $1\ ^3B_u$ state is of the order of 0.5 eV and drops to 0.32 eV in HCTD. Calculated vertical and adiabatic excitation energies of the T_1 states of all polyenes are presented in Table 9. While the destabilization of the ground-state energy is rather strong for the short members, the effect levels off for the long polyenes. States which exhibit large expansion coefficients for configurations with doubly occupied LUMO ($2\ ^1A_g$, $^1B_u^-$, $3\ ^1A_g$, $3\ ^3B_u$) experience huge stabilization effects. Qualitatively this can be understood from the fact that the electron density in the LUMO is large for those bonds which correspond to short bonds in the T_1 state—at least in the central part of the chromophore. This stabilization of $(LUMO)^2$ occupations leads, for example, to the strange situation that the $2\ ^1A_g$ state drops below the $1\ ^3A_g$ state. Actually, the latter is slightly shifted to higher energies at the T_1 geometry. A further effect of the pronounced stabilization of the $^1B_u^-$ state is the significantly earlier crossover with the optically bright $^1B_u^+$ state which occurs already for $N = 6$ at the relaxed T_1 geometry. Inspite of its (HOMO → LUMO) character, the latter state is only slightly affected by the geometry relaxation in its triplet coupled counterpart. The reason for this behavior will become clear in the next section. Finally, for the longest polyenes ($N > 11$) we observe even a drop down of the $3\ ^1A_g$ state below the $^1B_u^+$ state. Although it will be more meaningful to check the order of states at the relaxed $^1B_u^+$ geometry (see following section), it is interesting to notice that a conical intersection between the two states occurs not far from the $^1B_u^+$ minimum.

3.2.3. Geometry Relaxation in the $^1B_u^+$ State. Although both the T_1 and the $^1B_u^+$ states are dominated by the (HOMO → LUMO) single excitation, relaxation of the nuclear coordinates leads to quite different equilibrium structures. We saw above that single and double bonds localize in the T_1 state in three different regions, i.e., the central and the two terminal parts of the polyene. These regions are separated by short sequences of C−C bonds where the bond alternation changes. In the $^1B_u^+$ state, double bond character is found for the terminal bonds, too, but in the central part of the polyene, bond lengths are nearly equalized (Figure 5). This bond length equalization is actually what one might have expected from a simple Walsh-type analysis of this electronic excitation. According to Walsh, the geometric parameters of a molecule correlate with the trends for the frontier orbital energies upon nuclear distortion.[81] The (N − 1) nodes of the HOMO are placed where the LUMO exhibits maximal amplitudes and vice versa. In the $^{1,3}B_u^+$ states, the HOMO and LUMO are singly occupied each and single and double bond character should thus level out.

The energy gain of the $^1B_u^+$ state upon geometry relaxation to the minimum is of the order of 0.3 eV in the short polyenes and about 0.2 eV in the long ones (Table 9). The energies of the $^1B_u^+$ states at the respective T_1 geometries are only slightly less favorable. Together these facts indicate that the $^1B_u^+$ potential energy hypersurfaces (PEHs) of the longer polyenes are rather flat with respect to synchronous, but antiphase distortions of neighboring C−C bond lengths. We note in passing that a similar observation is made for the electronic ground states of the long polyenes which mix in non-negligible amounts of (HOMO → LUMO)^2 character. Their absolute DFT/MRCI energies are nearly identical at the S_0 and $^1B_u^+$ minimum geometries. The data in Table 9 show that the difference between the adiabatic excitation energies of the 2 $^1A_g^-$ and $^1B_u^+$ state increases with the conjugation length and becomes nearly constant for the longer polyenes. Experimental energy differences that were corrected for solvent effects, have been published for octatetraene (0.791 eV), decapentaene (0.874 eV), and dodecahexaene (0.920 eV).[69] Our corresponding calculated values of 0.85, 0.93, and 0.97 eV are in the right ballpark and reflect the experimental trends. Also the calculated energy gap for $N = 11,\ 12,\ 13$ (roughly 1 eV) is in good agreement with the estimated long-chain limit of 7168 cm^{-1} (0.89 eV).[82]

Since there is no reversal of single- and double-bond character in the $^1B_u^+$ state, the energies of states that are characterized by a doubly occupied LUMO are less dramatically affected by the geometry relaxation than at the T_1 geometry. The stabilization of the $^1B_u^-$ state is sufficient, however, to cause a crossover between $N = 8$ and $N = 9$ where both states are near-degenerate. This near-degeneracy of states should manifest itself in perturbations in the spectra of these polyenes. It has to be noted, however, that the position of the intersection may vary with the solvent polarizability since the $^1B_u^+$ state is known to exhibit strong solvatochromic shifts. Actually, indications of such a near-degeneracy have been observed recently in femtosecond-resolved spectra of lutein, a pigment with 10 conjugated double bonds.[83] For carotenes with conjugation lengths $N \geq 9$, the presence of the $^1B_u^-$ state intermediate between $2^1A_g^-$ and $^1B_u^+$ has been postulated by several authors to assign resonance-Raman spectra or to explain the intricate relaxation dynamics of these compounds in femtosecond spectroscopy.[84−90]

3.2.4. Excited State Absorption. The longer polyenes show strong ESA and are thus interesting candidates for optical limiting. Triplet ESA will not be discussed here (although it is strong) because the triplet quantum yields of polyenes are known to be very small.

The excitation energies and oscillator strengths originating from the S_1 (2 ^1A_g) state are collected in Table 10 for both the ground-state and the relaxed excited-state geometries. The strong geometry dependence of the transition energy and intensity of the first ESA band (2 ^1A_g → 1 ^1B_u in short polyenes, 2 ^1A_g → 2 ^1B_u in the longer ones) is noteworthy. In particular for the long polyenes the use of the Franck−Condon approximation for the modeling of this ESA band appears questionable. The calculated ESA wavelengths of the longer polyenes are slightly smaller than half the laser wavelengths required for the 1 ^1A_g → 2 ^1A_g TPA, even if the systematic errors of the DFT/MRCI 2 ^1A_g excitation energies are taken into account. However, the energetic location of the upper $^1B_u^+$ state could be tuned into resonance by an appropriate choice of environment.

The very strong ESA in the visible and UVA range is due to (LUMO → LUMO + 1) excitations from S_1 to a higher, multiconfigurational ^1B_u state. In HT and OT, these excitations are spread over two valence ^1B_u states, but for the longer members of the series, the intensity is concentrated in a single transition. The numbering of the upper states may change upon geometry variation. At the relaxed S_1 geometry the upper state corresponds to 4 ^1B_u. While its transition frequency is less sensitive to geometry relaxation, we find a marked decrease of its oscillator strength. Energetically, it overlaps with the strong 1 ^1A_g → 1 $^1B_u^+$ one-photon absorption and will thus be difficult to observe in experiments with low time resolution.

Some polyenes fluoresce from the 1 $^1B_u^+$ (S_2/S_3) state. Therefore, ESA data have been computed for this state, too (Table 11). Huge oscillator strengths are found for an ESA transition in the visible to near-infrared spectral region. The upper ^1A_g state represents a (LUMO → LUMO + 1) excitation with respect to the 1 $^1B_u^+$ state. It drifts from 4 ^1A_g in HT to 6 ^1A_g in the longer polyenes, but retains its electronic structure. A similar observation was recently made by Mikhailov et al.[91] who employed SAC-CI and an a posteriori TDA method in combination with the B3-LYP functional.

Although the oscillator strengths for 1 $^1B_u^+$ → 6 ^1A_g are larger than for the primary 1 ^1A_g → 1 $^1B_u^+$ absorption, it will be difficult to reach a population inversion because the excitation energies do not match. Whether a tuning by solvent

Table 10. Singlet Excited State Absorption Bands ΔE_{ESA} [eV] of Linear Polyenes with Conjugation Length N[a]

	ΔE_{ESA} ($f(r)$)					
N	1 ^1B_u	2 ^1B_u	3 ^1B_u	4 ^1B_u	5 ^1B_u	6 ^1B_u
2 ^1A_g state, FC region						
3	0.13	1.24	3.32 (0.16)	3.98 (0.65)	5.01 (0.01)	5.99
4	0.31 (0.01)	1.26	3.38 (0.74)	3.64 (0.56)	3.65	4.67 (0.01)
5	0.42 (0.01)	1.15	2.52	3.14 (1.64)	3.36 (0.12)	3.40
6	0.51 (0.02)	1.04	2.52	2.89 (2.13)	3.07 (0.03)	3.18 (0.01)
7	0.57 (0.03)	0.94	2.42	2.67 (2.51)	2.77 (0.03)	3.00
8	0.61 (0.04)	0.86	2.27	2.48 (2.60)	2.51 (0.31)	2.84
9	0.64 (0.05)	0.79	2.14	2.29 (0.03)	2.34 (3.25)	2.70
10	0.66 (0.07)	0.72	2.00	2.10	2.21(3.62)	2.57
11	0.67	0.68 (0.09)	1.87	1.93	2.09 (3.97)	2.46
12	0.62	0.69 (0.11)	1.75	1.80	2.00 (4.30)	2.36
13	0.58	0.69 (0.14)	1.65	1.68	1.91 (4.60)	2.27

	1 ^1B_u	2 ^1B_u	3 ^1B_u	4 ^1B_u	5 ^1B_u	6 ^1B_u
2 ^1A_g state, relaxed[b]						
3	0.79(0.02)	1.46	3.97(0.30)	4.97(0.55)	5.38	6.62(0.01)
4	0.92(0.04)	1.33	3.83(0.97)	4.16(0.01)	4.58(0.37)	4.99
5	1.00(0.06)	1.20(0.01)	2.69	3.51(1.52)	3.96	4.26(0.24)
6	1.05(0.04)	1.09(0.06)	2.56	3.21(1.95)	3.60	4.00(0.18)
7	0.95	1.11(0.15)	2.43	2.97(2.29)	3.26	3.36
8	0.84	1.13(0.21)	2.25	2.75(2.56)	2.99	3.21
9	0.75	1.14(0.28)	2.08	2.56(2.82)	2.76	3.12
10	0.67	1.14(0.36)	1.92	2.41(3.02)	2.56	2.98
11	0.60	1.14(0.44)	1.77	2.28(3.17)	2.38	2.83
12	0.53	1.13(0.53)	1.64	2.16(3.31)	2.23	2.67
13	0.48	1.12(0.61)	1.52	2.06(3.41)	2.09(0.02)	2.52

[a] Oscillator strengths are displayed in parentheses. [b] DFT/MRCI excitation energies and oscillator strengths at the 1 ^3B_u minimum.

Table 11. Singlet Excited State Absorption Bands ΔE_{ESA} [eV] of Linear Polyenes with Conjugation Length N[a]

	ΔE_{ESA} ($f(r)$)			
N	3 ^1A_g	4 ^1A_g	5 ^1A_g	6 ^1A_g
1 ^1B_u state for $N \le 10$, 2 ^1B_u for $N \ge 11$, FC region				
3	2.09	2.94 (1.11)	3.12 (0.95)	4.86
4	1.67	1.82	2.45 (0.85)	2.64 (2.02)
5	1.56	1.60 (0.01)	2.02 (0.35)	2.30 (3.28)
6	1.36 (0.01)	1.42(0.01)	1.64 (0.13)	2.02 (4.22)
7	1.16 (0.01)	1.27 (0.02)	1.34 (0.07)	1.81 (4.96)
8	0.99 (0.01)	1.10 (0.06)	1.16	1.65 (5.66)
9	0.83	0.90 (0.05)	1.06	1.50 (6.31)
10	0.70	0.75(0.04)	0.98 (0.01)	1.39 (6.92)
11	0.59	0.63(0.04)	0.91 (0.01)	1.29 (7.26)
12	0.50	0.53(0.03)	0.85	1.21 (8.12)
13	0.42	0.44 (0.03)	0.79 (0.01)	1.14 (8.67)

	3 ^1A_g	4 ^1A_g	5 ^1A_g	6 ^1A_g
1 ^1B_u state for $N \le 7$, 2 ^1B_u for $N \ge 8$, relaxed				
3	2.23(0.01)	2.58(1.50)	2.91(0.47)	5.11
4	1.41(0.01)	1.93(0.03)	2.27(1.88)	2.44(0.89)
5	1.28(0.02)	1.69(0.04)	1.89(0.36)	2.06(3.16)
6	1.08(0.02)	1.51(0.06)	1.52(0.09)	1.82(4.13)
7	0.90(0.02)	1.23(0.04)	1.36(0.05)	1.63(4.86)
8	0.72(0.02)	1.00(0.02)	1.24(0.05)	1.48(5.25)
9	0.57(0.02)	0.81(0.01)	1.13(0.05)	1.35(6.22)
10	0.45(0.01)	0.66(0.01)	1.05(0.06)	1.24(6.84)
11	0.34(0.01)	0.53	0.97(0.05)	1.15(7.44)
12	0.26(0.01)	0.44	0.92(0.07)	1.08(8.01)
13	0.19(0.01)	0.35	0.86(0.05)	1.01(8.57)

[a] Oscillator strengths are displayed in parentheses.

effects is possible here cannot be concluded from the present data since the behavior of the upper state is not known.

Conclusions

The purpose of the present theoretical study was a critical assessment of the DFT/MRCI method in notoriously difficult cases where TDDFT in combination with standard functionals fails. In addition, the effects of various technical parameters of the calculations on the properties of the electronic states were thoroughly investigated.

The first of these cases is the position of the ^1L_a (HOMO−LUMO) transition in polyacenes. The DFT/MRCI energies follow the experimental trends with a tendency of slightly underestimating the true ^1L_a and ^1L_b excitation energies. Their accuracy is comparable to the one of the considerably more expensive ab initio CC2 method which overestimates the experimental values by approximately the same amount.

Static electron correlation in the multiconfigurational 2 ^1A_g state of linear polyenes and α,ω-diphenyl-polyenes is typically severely underrated by single-reference response methods such as TDDFT and CC2 with the result that the order of the 2 ^1A_g and 1 ^1B_u states is reversed in these calculations. The energy gap between these states is reproduced correctly by the DFT/MRCI method, but again the absolute excitation energies are somewhat too low. In linear polyenes with conjugation lengths 8 to 11 an interesting phenomenon is observed in the calculations. Upon geometry relaxation in the primarily excited $^1B_u^+$ state the multiconfigurational $^1B_u^-$ state is stabilized to an extent that an intersection between the corresponding potential energy hypersurfaces takes place. In polyenes with conjugation lengths 12 and 13, the $^1B_u^-$ state represents the S_2 state already in the Franck−Condon region. In the latter compounds, even the 3 ^1A_g drops down below the $^1B_u^+$ state. The presence of conical intersections between the optically bright $^1B_u^+$ state and dark singlet states lays the ground for the supposition that internal conversion after photoexcitation is extremely fast in the longer polyenes. Excited state absorption wave-

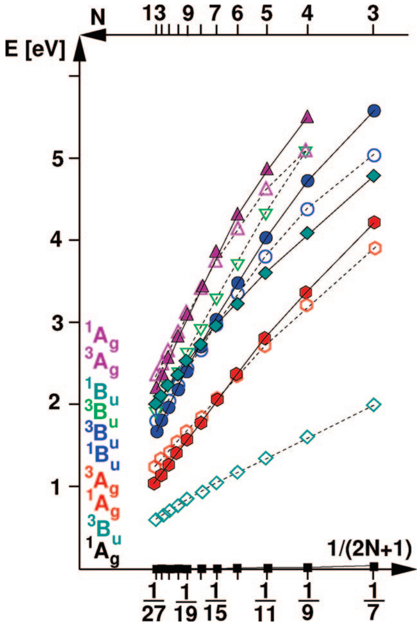


Figure 8. DFT/MRCI vertical electronic excitation energies of *all-trans*-polyenes at the minimum geometry of the optically bright ^1B_u (HOMO → LUMO) state as functions of the conjugation length *N*. The DFT/MRCI energy of the respective 1 ^1A_g ground-state minimum has been chosen as energy offset. For an explanation of symbols, see Figure 6.

lengths and intensities are found to depend strongly on the nuclear geometry. In the gas phase, the wavelengths for ESA from the first excited singlet to the $^1B_u^+$ do not match with the laser wavelengths used for its two-photon excitation from the ground state. Due to its solvatochromism, the energetic location of the upper state could be tuned into resonance by an appropriate choice of environment, however.

With regard to the treatment of large π-systems it is encouraging that the reduction of the basis set quality from TZVPP over TZVP to SV(P) has almost no effect on the excitation energies and oscillator strengths. A more critical parameter is the ground-state equilibrium nuclear arrangement, at least in case of the polyene chains. As exemplified for *trans*,*trans*-1,3,5,7-octatetraene, where an experimental crystal structure is known, only the Hartree−Fock method yields C−C bond distances close to the experimental ones, most certainly due to a fortuitous error cancelation. All correlated methods applied (including the density functionals B-LYP, B3-LYP, and BH-LYP as well as the RIMP2, MP2, and CASSCF wave function approaches) underestimate the bond length alternation in the central part of the chromophore. These geometric deficiencies introduce a bias in favor of the electronically excited states of the polyenes, since the C−C bond length alternation is less pronounced or even reversed in the latter, offering an explanation for the fact that the DFT/MRCI electronic excitation energies of linear polyenes and α, ω-diphenyl-polyenes are consistently too low while the energy gaps between the excited states are reproduced well.

Acknowledgment. Intensive discussions with Alfred Holzwarth (MPI Mülheim) and Andreas Dreuw (Universität Frankfurt) during the progress of this work have been extremely valuable. Financial support from the Deutsche Forschungsgemeinschaft and the Heinrich-Heine-Universität Düsseldorf in the framework of the “Sonderforschungsbereich 663” is gratefully acknowledged.

Supporting Information Available: Tables S1 and S2 and Figures S1−S8. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Polivka, T.; Sundström, V. *Chem. Rev.* **2004**, *104*, 2021.

(2) Hsu, C.-P.; Hirata, S.; Head-Gordon, M. *J. Phys. Chem.* **2001**, *105*, 451.

(3) Wanko, M.; Garavelli, M.; Bernardi, F.; Niehaus, T. A.; Frauenheim, T.; Elstner, M. *J. Chem. Phys.* **2004**, *120*, 1674.

(4) Catalán, J.; de Paz, J. L. G. *J. Chem. Phys.* **2006**, *124*, 034306.

(5) Dreuw, A.; Head-Gordon, M. *Chem. Rev.* **2005**, *105*, 4009.

(6) Starcke, J. H.; Wormit, M.; Schirmer, J.; Dreuw, A. *Chem. Phys.* **2006**, *329*, 39.

(7) Hudson, B. S.; Kohler, B. E. *Chem. Phys. Lett.* **1972**, *14*, 299.

(8) Hudson, B. S.; Kohler, B. E. *J. Chem. Phys.* **1973**, *59*, 4984.

(9) Schulten, K.; Karplus, M. *Chem. Phys. Lett.* **1972**, *14*, 305.

(10) Tavan, P.; Schulten, K. *J. Chem. Phys.* **1979**, *70*, 5407.

(11) Tavan, P.; Schulten, K. *Phys. Rev. B* **1987**, *36*, 4337.

(12) Serrano-Andrés, A.; Lindh, R.; Roos, B. O.; Merchán, M. *J. Phys. Chem.* **193**, *97*, 9360.

(13) Serrano-Andrés, A.; Merchán, M.; Nebot-Gil, I.; Lindh, R.; Roos, B. O. *J. Chem. Phys.* **1993**, *98*, 3151.

(14) Nakayama, K.; Nakano, H.; Hirao, K. *Int. J. Quantum Chem.* **1998**, *66*, 157.

(15) Luo, Y.; Ågren, H.; Stafström, S. *J. Chem. Phys.* **1994**, *98*, 7782.

(16) Cronstrand, P.; Christiansen, O.; Norman, P.; Ågren, H. *Phys. Chem. Chem. Phys.* **2001**, *3*, 2567.

(17) Schreiber, M.; Silva-Junior, M. R.; Thiel, W.; Sauer, S. P. A. *J. Chem. Phys.* **2008**, *128*, 134110.

(18) Schirmer, J. *Phys. Rev. A* **1982**, *26*, 2395.

(19) Parac, M.; Grimme, S. *Chem. Phys.* **2003**, *292*, 11.

(20) Roos, B. O.; Andersson, K.; Fülscher, M. P. *Chem. Phys. Lett.* **1992**, *192*, 5.

(21) Grimme, S.; Waletzke, M. *J. Chem. Phys.* **1999**, *111*, 5645.

(22) Marian, C. M. *J. Chem. Phys.* **2005**, *122*, 104314.

(23) Kleinschmidt, M.; Tatchen, J.; Marian, C. M. *J. Chem. Phys.* **2006**, *124*, 124101.

(24) Tatchen, J.; Marian, C. M. *Phys. Chem. Chem. Phys.* **2006**, *8*, 2133.

(25) Marian, C. M. *J. Phys. Chem. A* **2007**, *111*, 1545.

(26) Perun, S.; Tatchen, J.; Marian, C. M. *ChemPhysChem* **2008**, *9*, 282.

(27) Salzmann, S.; Tatchen, J.; Marian, C. M. *J. Photochem. Photobiol. A: Chem.* **2008**, *198*, 221.

(28) Schäfer, A.; Huber, C.; Ahlrichs, R. *J. Chem. Phys.* **1994**, *100*, 5829.

(29) Furche, F.; Ahlrichs, R. *J. Chem. Phys.* **2002**, *117*, 7433.

(30) Ahlrichs, R.; et al. *TURBOMOLE*; version 5.6, Universität Karlsruhe, 2002.

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

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

(33) Vahtras, O.; Almlöf, J.; Feyereisen, M. W. *Chem. Phys. Lett.* **1993**, *213*, 514.

(34) Eichkorn, K.; Treutler, O.; Öhm, H.; Häser, M.; Ahlrichs, R. *Chem. Phys. Lett.* **1995**, *240*, 283.

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

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

(37) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372.

(38) Biermann, D.; Schmidt, W. *J. Am. Chem. Soc.* **1980**, *102*, 3163.

(39) Gavin, R. M., Jr.; Rice, S. A. *J. Chem. Phys.* **1974**, *60*, 3231.

(40) Gavin, R. M., Jr.; Weisman, C.; Rice, S. A. *J. Chem. Phys.* **1978**, *68*, 522.

(41) Leopold, D. G.; Vaida, V.; Granville, M. F. *J. Chem. Phys.* **1984**, *81*, 4210.

(42) Leopold, D. G.; Pendley, R. D.; Roebber, J. L.; Hemley, R. J.; Vaida, V. *J. Chem. Phys.* **1984**, *81*, 4218.

(43) Heimbrook, L. A.; Kohler, B. E.; Levy, I. J. *J. Chem. Phys.* **1984**, *81*, 1592.

(44) Buma, W. J.; Kohler, B. E.; Song, K. *J. Chem. Phys.* **1991**, *94*, 6367.

(45) Petek, H.; Bell, A. J.; Christensen, R. L.; Yoshihara, K. *J. Chem. Phys.* **1992**, *96*, 2412.

(46) McDiarmid, R. *Adv. Chem. Phys.* **1999**, *110*, 177.

(47) Pfanstiel, J. F.; Pratt, D. W.; Tounge, B. A.; Christensen, R. L. *J. Phys. Chem. A* **1999**, *103*, 2337.

(48) Post, D. E.; Hetherington, W. M., III; Hudson, B. *Chem. Phys. Lett.* **1975**, *35*, 259.

(49) Flicker, W. M.; Mosher, O. A.; Kuppermann, A. *Chem. Phys. Lett.* **1977**, *45*, 492.

(50) Frueholz, R. P.; Kuppermann, A. *J. Chem. Phys.* **1978**, *69*, 3433.

(51) Allan, M.; Neuhaus, L.; Haselbach, E. *Helv. Chim. Acta* **1984**, *67*, 1776.

(52) Trætteberg, M. *Acta Chem. Scand.* **1968**, *22*, 628.

(53) Granville, M. F.; Holtom, G. R.; Kohler, B. E. *J. Chem. Phys.* **1980**, *72*, 4671.

(54) Kohler, B. E.; Terpougov, V. *J. Chem. Phys.* **1996**, *104*, 9297.

(55) Kohler, B. E.; Terpougov, V. *J. Chem. Phys.* **1998**, *108*, 9586.

(56) Cave, R. J.; Davidson, E. R. *J. Phys. Chem.* **1988**, *92*, 614.

(57) Cave, R. J.; Davidson, E. R. *Chem. Phys. Lett.* **1988**, *148*, 190.

(58) Cave, R. J.; Davidson, E. R. *J. Phys. Chem.* **1988**, *92*, 2173.

(59) Li, X.; Paldus, J. *Int. J. Quantum Chem.* **1999**, *74*, 177.

(60) Woywod, C.; Livingood, W. C.; Frederick, J. H. *J. Chem. Phys.* **2000**, *112*, 613.

(61) Woywod, C. *Chem. Phys.* **2005**, *311*, 321.

(62) Petrenko, T.; Neese, F. *J. Chem. Phys.* **2007**, *127*, 164319.

(63) Cave, R. J.; Zhang, F.; Maitra, N. T.; Burke, K. *Chem. Phys. Lett.* **2004**, *389*, 39.

(64) Fujii, T.; Kamata, A.; Shimizu, M.; Adachi, Y.; Meada, S. *Chem. Phys. Lett.* **1985**, *115*, 369.

(65) Gavin, R. M., Jr.; Risenberg, S.; Rice, S. A. *J. Chem. Phys.* **1973**, *58*, 3160.

(66) Myers, A. B.; Pranata, K. S. *J. Phys. Chem.* **1989**, *93*, 5079.

(67) Davidson, E. R.; Jarzecki, A. A. *Chem. Phys. Lett.* **1998**, *285*, 155.

(68) Baughman, R. H.; Kohler, B. E.; Levy, I. J.; Spangler, C. *Synth. Met.* **1985**, *11*, 37.

(69) D'Amico, K. L.; Manos, C.; Christensen, F. L. *J. Am. Chem. Soc.* **1980**, *102*, 1777.

(70) Fang, H. L.-B.; Thrash, R. J.; Leroi, G. E. *J. Chem. Phys.* **1977**, *67*, 3389.

(71) Fang, H. L.-B.; Thrash, R. J.; Leroi, G. E. *Chem. Phys. Lett.* **1978**, *57*, 59.

(72) Horwitz, J. S.; Itoh, T.; Kohler, B. E.; Spangler, C. W. *J. Chem. Phys.* **1987**, *87*, 2433.

(73) Kohler, B. E.; Itoh, T. *J. Chem. Phys.* **1988**, *92*, 5120.

(74) Weiss, V.; Port, H.; Wolf, H. C. *Chem. Phys. Lett.* **1992**, *192*, 289.

(75) Pariser, R. *J. Chem. Phys.* **1956**, *24*, 324.

(76) Kuki, M.; Koyama, Y.; Nagae, H. *J. Phys. Chem.* **1991**, *95*, 7171.

(77) Takahashi, O.; Watanabe, M.; Kikuchi, O. *Theochem—J. Mol. Struct.* **1999**, *469*, 121.

(78) Takahashi, O.; Watanabe, M.; Kikuchi, O. *Int. J. Quantum Chem.* **1998**, *67*, 101.

(79) Ma, H.; Liu, C.; Jiang, Y. *J. Chem. Phys.* **2004**, *120*, 9316.

(80) Ma, H.; Cai, F.; Liu, C.; Jiang, Y. *J. Chem. Phys.* **2005**, *122*, 104909.

(81) Walsh, A. D. *J. Chem. Soc.* **1953**, 2260; 2266; 2288; 2296; 2301; 2306; 2318; 2321; 2325; 2330.

(82) Andersson, P. O.; Gillbro, T. *J. Chem. Phys.* **1995**, *103*, 2509.

(83) Holzwarth, A. L. in preparation, 2008.

(84) Sashima, T.; Koyama, Y.; Yamada, T.; Hashimoto, H. *J. Phys. Chem. B* **2000**, *104*, 5011.

(85) Yoshizawa, M.; Aoki, H.; Hashimoto, H. *Bull. Chem. Soc. Jpn.* **2002**, *75*, 949.

(86) Yoshizawa, M.; Aoki, H.; Hashimoto, H. *Phys. Rev. B* **2001**, *63*, 180301.

(87) Yoshizawa, M.; Aoki, H.; Ue, M.; Hashimoto, H. *Phys. Rev. B* **2003**, *67*, 174302.

(88) Larsen, D.; Papagiannakis, E.; I.H.M. van Stokkum, M. V.; Kennis, J.; van Grondelle, R. *Chem. Phys. Lett.* **2003**, *381*, 733.

(89) Polli, D.; Cerullo, G.; Lanzani, G.; Silvestri, S. D.; Yanagi, K.; Hashimoto, H.; Cogdell, R. *Phys. Rev. Lett.* **2004**, *93*, 163002.

(90) Kosumi, D.; Yanagi, K.; Nishio, T.; Hashimoto, H.; Yoshizawa, M. *Chem. Phys. Lett.* **2005**, *408*, 89.

(91) Mikhailov, I. A.; Tafur, S.; Masunov, A. E. *Phys. Rev. A* **2008**, *77*, 012510.

(92) Kohler, B. E.; Terpougov, V. *J. Chem. Phys.* **198**, *108*, 9586.

CT8001738

JCTC Journal of Chemical Theory and Computation

Coupling Accelerated Molecular Dynamics Methods with Thermodynamic Integration Simulations

César Augusto F. de Oliveira,* Donald Hamelberg, and J. Andrew McCammon

*Howard Hughes Medical Institute, Center for Theoretical Biological Physics,
Department of Chemistry and Biochemistry and Department of Pharmacology,
University of California at San Diego, La Jolla, California 92093-0365*

Received May 12, 2008

Abstract: In this work we propose a straightforward and efficient approach to improve accuracy and convergence of free energy simulations in condensed-phase systems. We also introduce a new accelerated Molecular Dynamics (MD) approach in which molecular conformational transitions are accelerated by lowering the energy barriers while the potential surfaces near the minima are left unchanged. All free energy calculations were performed on the propane-to-propane model system. The accuracy of free energy simulations was significantly improved when sampling of internal degrees of freedom of solute was enhanced. However, accurate and converged results were only achieved when the solvent interactions were taken into account in the accelerated MD approaches. The analysis of the distribution of boost potential along the free energy simulations showed that the new accelerated MD approach samples efficiently both low- and high-energy regions of the potential surface. Since this approach also maintains substantial populations in regions near the minima, the statistics are not compromised in the thermodynamic integration calculations, and, as a result, the ensemble average can be recovered.

Introduction

Free energy is probably the most important quantity in thermodynamics and one of the central topics in biophysics.[1,2] Nevertheless, for many relevant systems with local minimum energy configurations separated by energy barriers, efficient and accurate calculation of this property is still a big challenge in computational chemistry. Free energy differences between different states can be calculated through Free Energy Perturbation (FEP) and Thermodynamic Integration (TI) methods.[3−10] Since the first application of the methodology to the calculation of the relative free energies of ligand binding and solvation of the organic molecules methanol and ethane,[1,8] FEP and TI have been widely used to study a wide range of processes such as solvation, phase transitions, ligand binding, and protein−protein interactions, just to name a few.[5,11−14] These methods, which are firmly rooted in statistical mechanics, are usually combined with molecular dynamics (MD) or Monte Carlo (MC) simulations.[15,16]

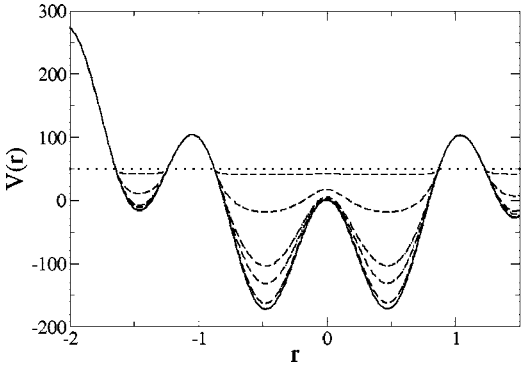


Figure 1. Schematic representation of a hypothetical true (solid line) and modified (dashed line) potential energy function with different values of α. The modified potential (generated with eq 1) converges to the true potential at large values of α. The dotted line corresponds to the boost energy *E*.

The data obtained from these simulations allows us to quantitatively evaluate free energy changes and understand, at molecular level, the structural and energetic factors governing the process.

* Corresponding author e-mail: cesar@mccammon.ucsd.edu.

10.1021/ct800160q CCC: $40.75
Published on Web 08/13/2008

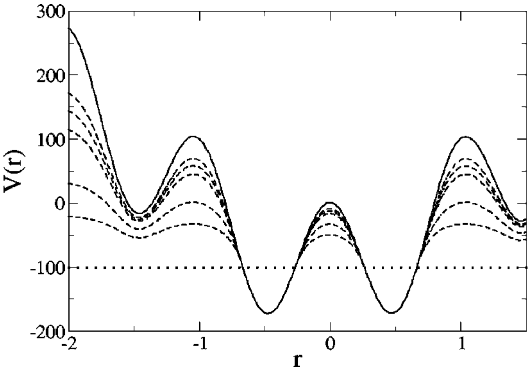


Figure 2. Schematic representation of a hypothetical true (solid line) and modified (dashed line) potential energy function with different values of α. The modified potential (generated with eq 3) converges to the true potential at large values of α. The dotted line corresponds to the boost energy *E*.

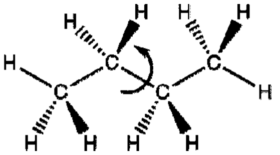


Figure 3. Butane molecule.

However, to obtain accurate free energy values, major issues like free energy convergence and conformational sampling still need to be addressed. Although these topics have been mainly discussed as independent issues, convergence and the amount of sampling are strictly connected.[17] For instance, for processes that involve large conformational changes and reorganization of solvent, poor sampling can trap the system in local minima and, as a consequence, lead to apparent but false convergence. In other words, in these cases the calculated free energy might correspond to pseudoconverged values obtained from trapped local conformations. As we will show in this paper, even for a very simple system, like propane-to-propane transformation, independent free energy calculations carried out with conventional MD simulation may not be able to reproduce accurately the correct free energy difference, though the simulations may show apparently converged values. Quantitative prediction of free energy change is only obtained when configuration sampling is efficiently improved.

A large number of techniques have been introduced to enhance sampling over configuration space.[18−32] A straightforward way of modifying the potential energy surface to enhance sampling has been proposed by Hamelberg et al.[33] This approach, which is based on earlier work of Voter,[34,35] has proved to be efficient in accelerating not only conformational transitions[36−39] but also millisecond time scale motions of a protein in explicit water.[40]

In this work we propose a simple and efficient approach to improve accuracy and convergence of free energy simulations in condensed-phase systems. The main idea is to integrate the accelerated MD approach with free energy simulations. Although the formulation and the results presented here were obtained by coupling TI with the accelerated MD method (aMD), the procedure can be easily extended to the FEP approach. To check convergence and accuracy of the TI simulations, all calculations were performed on the propane-to-propane system. This system was chosen because i) the correct free energy result is rigorously equal to zero and ii) similar "zero-free energy change" systems have been used before as model systems to compare the efficiency and convergence of different approaches to free energy calculations.[41,42]

Theory

In order to enhance sampling by increasing the escape rate from potential energy wells, the accelerated MD approach modifies the energy landscape by adding a boost potential, $\Delta V(r)$, to the original potential surface every time $V(r)$ is below a predefined energy level E (Figure 1). In other words,

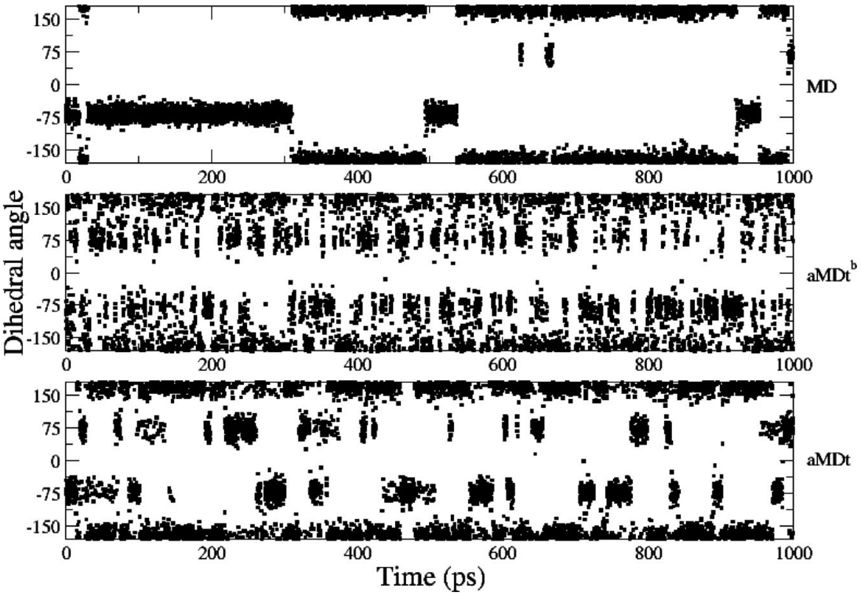


Figure 4. Plots of dihedral angle of the butane molecule, as defined in Figure 3, sampled with normal MD, aMDt^b, and aMDt approaches.

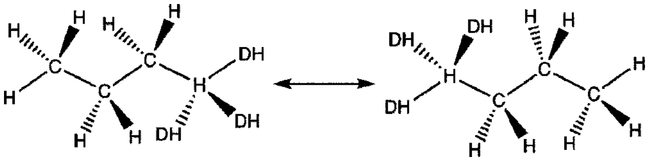


Figure 5. Propane-to-propane transformation. DH stands for dummy atoms.

$V^*(r) = V(r) + \Delta V(r)$. In the Hamelberg et al.[33] implementation, $\Delta V(r)$ is given by

$$\Delta V(r) = \begin{cases} 0, & V(r) \geq E \\ \dfrac{(E - V(r))^2}{\alpha + (E - V(r))}, & V(r) < E \end{cases} \tag{1}$$

where α modulates the depth and the local roughness of the energy basins in the modified potential. Since the torsional potential governs the rate of sampling of biomolecular rotameric states, the boost potential has been largely applied to the torsional term of the potential energy function. This approach, which will be referred to as aMDt, has been successfully applied to study several biological systems and processes.[37−39,43,44]

More recently, Hamelberg et al. introduced a dual boost approach in order to efficiently sample both the torsional degrees of freedom and the diffusive motions.[36] In this implementation, two boost potentials are applied separately to the potential energy. While the first one is applied only to the torsional terms, the second one is added to the total potential energy (aMDtT). The modified potential is given by

$$V^*(r) = \{V_0(r) + [V_t(r) + \Delta V_t(r)]\} + \Delta V_T(r) \tag{2}$$

where $\Delta V_t(r)$ and $\Delta V_T(r)$ are the boost potentials applied to the torsional terms $V_t(r)$ and the total potential $V_T(r)$. $V_0(r)$ is the potential energy excluding contribution from torsional terms. Both boost potentials are defined according to eq 1. Here $V_T(r)$ is defined as $V_T(r) = V_0(r) + V_t(r) + \Delta V_t(r)$.

The correct canonical averages of an observable, calculated from configurations sampled on the modified potential energy surface, is then fully recovered from the accelerated MD simulations by reweighting each point in the configuration space by $\exp\{\beta[\Delta V(r)]\}$. In the dual boost approach, the boost factor is given by $\exp\{\beta[\Delta V_t(r)+\Delta V_T(r)]\}$.

New Accelerated MD Approach. In this work, a third approach is introduced in which molecular conformational transitions are accelerated by lowering the energy barriers, while the potential surfaces near the minima are left unchanged. The idea behind this approach has been used before by Darve et al. to calculate free energies by applying a scale-force molecular dynamics algorithm.[27]

Owing to the symmetry of eq 1 in relation to E and $V(r)$, this approach can be easily implemented by simply redefining eq 1 as

$$\Delta V(r) = \begin{cases} \dfrac{(V(r) - E)^2}{\alpha + (V(r) - E)}, & V(r) \geq E \\ 0, & V(r) < E \end{cases} \tag{3}$$

In this implementation, the boost potential, $\Delta V(r)$, is subtracted from the true potential $V(r)$ whenever the potential

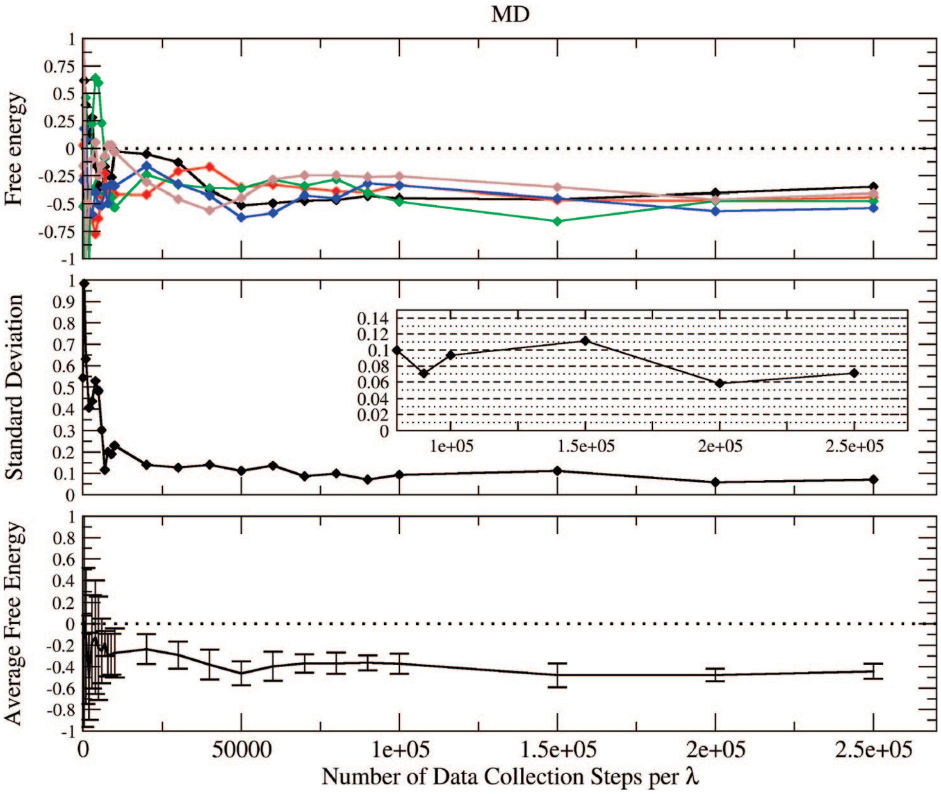


Figure 6. Free energy change in kcal/mol, calculated for the propane-to-propane simulations as a function of time from five independent simulations (top). Standard deviation of the results from the five independent simulations (middle). The inset plot shows in detail the standard deviation as a function of time for the points with at least 80 × 10^3 of data collection steps per λ. The same number of equilibration and data collection steps were used for each λ. Average free energy change and the error associated with each point were calculated from the five independent simulations (bottom). Units are in kcal/mol.

$V(r)$ is greater than the boost energy E; in this case, the simulation is performed on the modified potential $V^*(r) = V(r) - \Delta V(r)$. On the other hand, when $V(r)$ is below the energy level E, the simulation is performed on the true potential $V^*(r) = V(r)$. Hereafter, this approach will be referred to as aMD^b and aMDt^b when applied to the torsional terms of the potential energy. Figures 1 and 2 illustrate a schematic representation of a hypothetical one-dimensional potential modified using eqs 1 and 3, respectively. In both cases, as α decreases, the modified potential becomes flatter, and as α increases, the modified landscape asymptotically approaches the unmodified potential.

In this approach, analogously to Hamelberg et al.'s implementation, the correct canonical ensemble averages of an observable are fully recovered by reweighting each configuration by the Boltzmann factor of the negative of the boost potential energy, $\exp\{-\beta[\Delta V(r)]\}$. The application of this schema into the dual boost approach is straightforward. In this case, eq 2 is simply redefined as $V^*(r) = \{V_0(r)+[V_t(r)-\Delta V_t(r)]\}-\Delta V_T(r)$, and the boost factor as $\exp\{-\beta[\Delta V_t(r)+\Delta V_T(r)]\}$. This implementation will be referred to as aMDtT^b.

Coupling Accelerated MD Approach with Thermodynamic Integration Simulations. Thermodynamic integration is a commonly used technique to compute the difference in free energy between two thermodynamic states, which differ from each other according to their intermolecular or intramolecular interaction potentials.[3,8,12] In this case, the interaction potential can be expressed as a function of a coupling parameter, λ, that determines the state of the system.[10] Thus, by defining the free energy, F, as a continuous function of λ, the difference in free energy between two states is given by

$$\Delta F = \int_{\lambda=0}^{\lambda=1} \frac{\partial F(\lambda)}{\partial \lambda} d\lambda \tag{4}$$

where $\lambda = 0$ and 1 correspond to the initial and final states, respectively. Since F(λ) can be written as

$$F(\lambda) = -k_b T \ln Q(\lambda) \tag{5}$$

ΔF can be rewritten as[45]

$$\Delta F = \int -\beta \frac{1}{Q(\lambda)} \frac{\partial Q(\lambda)}{\partial \lambda} d\lambda \tag{6}$$

where Q is the partition function of the system, $\beta = 1/k_bT$, k_b is Boltzmann's constant, and T is the temperature. Here, we use the partition function for canonical ensemble, which is defined as[46]

$$Q_{NVT} = \frac{1}{N!}\frac{1}{h^{3N}} \int\int drdp \exp[-\beta H(p,r)] \tag{7}$$

where N is the number of particles, h is Planck's constant, p and r are the momenta and positions of the particles, and H is Hamiltonian of the system. Substituting eq 7 into eq 6 and deriving in respect to λ,[45] we obtain

$$\frac{\partial F(\lambda)}{\partial \lambda} = \frac{\int\int dpdr \frac{\partial \mathrm{H}(p,r)}{\partial \lambda} \exp[-\beta \mathrm{H}(p,r)]}{\int\int dpdr \exp[-\beta \mathrm{H}(p,r)]} \tag{8}$$

Assuming that the kinetic energy term is separable and not dependent on λ, eq 8 can be rewritten in terms of the potential energy $V(r)$ of the system

$$\frac{\partial F(\lambda)}{\partial \lambda} = \frac{\int dr \frac{\partial V(r)}{\partial \lambda} \exp[-\beta V(r)]}{\int dr \exp[-\beta V(r)]} \tag{9}$$

and, finally

$$\Delta F = \int_{\lambda=0}^{\lambda=1} \left\langle \frac{\partial V(r,\lambda)}{\partial \lambda} \right\rangle_{\lambda} d\lambda \tag{10}$$

where the integrand is the ensemble average of $\partial V/\partial \lambda$ calculated on the original potential $V(r)$ at a specific value of λ, and ΔF is the free energy difference between the initial ($\lambda = 0$) and final ($\lambda = 1$) states obtained on the unmodified potential surface, $V(r)$.

Similarly, for the modified potential we have

$$\frac{\partial F(\lambda)_*}{\partial \lambda} = \frac{\int dr \frac{\partial V(r)}{\partial \lambda} \exp[-\beta V^*(r)]}{\int dr \exp[-\beta V^*(r)]} \tag{11}$$

now, the ensemble average of true $\partial V(r)/\partial \lambda$ is performed over the modified potential $V^*(r)$.

Since both approaches can be coupled with TI simulations, we will first express $V^*(r)$ as

$$V^*(r) = V(r) + \Delta V \tag{12}$$

In this case, eq 11 can be rewritten as

$$\begin{aligned}\frac{\partial F(\lambda)_*}{\partial \lambda} &= \frac{\int dr \frac{\partial V(r)}{\partial \lambda} \exp\{-\beta[V(r) + \Delta V]\}}{\int dr \exp\{-\beta[V(r) + \Delta V]\}} \\ &= \frac{\int dr \frac{\partial V(r)}{\partial \lambda} \exp[-\beta V(r)] \exp[-\beta \Delta V]}{\int dr \exp[-\beta V(r)] \exp[-\beta \Delta V]}\end{aligned} \tag{13}$$

The Boltzmann distribution can be extracted from the non-Boltzmann distribution using the method introduced by Torrie et al.[19] The corrected canonical distribution can then be recovered by reweighting the phase space of the modified potential by multiplying the integrand by the strength of the bias at each position, which in this case corresponds to $\exp[\beta \Delta V]$.

$$\frac{\partial F(\lambda)_C}{\partial \lambda} = \frac{\int dr \frac{\partial V(r)}{\partial \lambda} \exp[-\beta V(r)] \exp[-\beta \Delta V] \exp[\beta \Delta V]}{\int dr \exp[-\beta V(r)] \exp[-\beta \Delta V] \exp[\beta \Delta V]} = \frac{\partial F(\lambda)}{\partial \lambda} \tag{14}$$

Thus, the corrected ensemble average of ΔF^C can be obtained by dividing both the numerator and the denominator of eq 11 by $\int dr \exp[-\beta V(r)] \exp[-\beta \Delta V] = \int dr \exp[-\beta V^*(r)]$

$$\frac{\partial F(\lambda)_C}{\partial \lambda} = \frac{\int dr \frac{\partial V(r)}{\partial \lambda} \exp[-\beta V(r)] \exp[-\beta \Delta V] \exp[\beta \Delta V] / \int dr \exp[-\beta V(r)] \exp[-\beta \Delta V]}{\int dr \exp[-\beta V(r)] \exp[-\beta \Delta V] \exp[\beta \Delta V] / \int dr \exp[-\beta V(r)] \exp[-\beta \Delta V]}$$

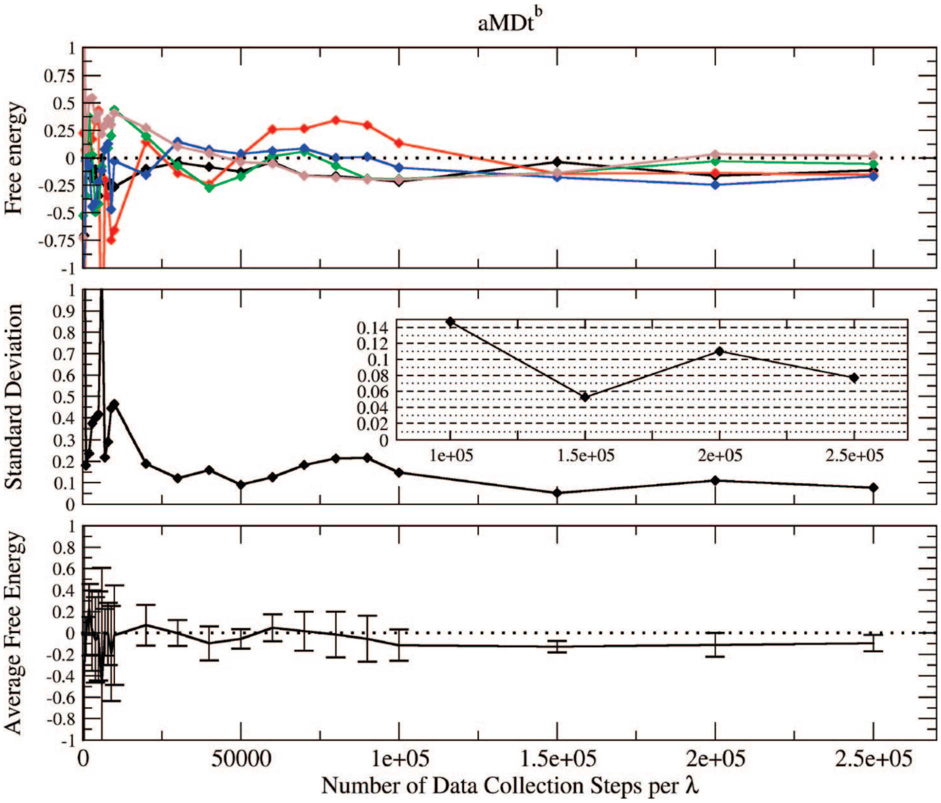


Figure 7. Free energy change in kcal/mol, calculated for the propane-to-propane simulations as a function of time from five independent simulations (top). Standard deviation of the results from the five independent simulations (middle). The inset plot shows in detail the standard deviation as a function of time for the points with at least 80 × 10^3 of data collection steps per λ. The same number of equilibration and data collection steps were used for each λ. Average free energy change and the error associated with each point were calculated from the five independent simulations (bottom). Units are in kcal/mol.

$$= \frac{\int dr \frac{\partial V(r)}{\partial \lambda} \exp[-\beta V^*(r)] \exp[\beta \Delta V] / \int dr \exp[-\beta V^*(r)]}{\int dr \exp[-\beta V^*(r)] \exp[\beta \Delta V] / \int dr \exp[-\beta V^*(r)]}$$

and integrating over λ.

$$\begin{aligned} \Delta F^C &= \int_{\lambda=0}^{\lambda=1} \left[\left\langle \frac{\partial V(r,\lambda)}{\partial \lambda} \exp[\beta \Delta V] \right\rangle_{\lambda *} \middle/ \langle \exp[\beta \Delta V] \rangle_{\lambda *} \right] d\lambda \\ &= \int_{\lambda=0}^{\lambda=1} \left\langle \frac{\partial V(r,\lambda)}{\partial \lambda} \right\rangle_{\lambda} d\lambda \\ &= \Delta F \end{aligned} \tag{16}$$

Analogously, if the modified potential is defined as $V^*(r) = V(r) - \Delta V$, eq 16 can be redefined as

$$\begin{aligned} \Delta F^C &= \int_{\lambda=0}^{\lambda=1} \left[\left\langle \frac{\partial V(r,\lambda)}{\partial \lambda} \exp[-\beta \Delta V] \right\rangle_{\lambda *} \middle/ \langle \exp[-\beta \Delta V] \rangle_{\lambda *} \right] d\lambda \\ &= \int_{\lambda=0}^{\lambda=1} \left\langle \frac{\partial V(r,\lambda)}{\partial \lambda} \right\rangle_{\lambda} d\lambda \\ &= \Delta F \end{aligned} \tag{17}$$

Therefore, independent of the approach applied, the accelerated molecular dynamics simulation method converges to the canonical distribution, and the corrected canonical ensemble average of the system is obtained by simply reweighting each point in the configuration phase space on the modified potential by the strength of the Boltzmann factor of the bias energy, $\exp[\beta \Delta V]$ or $\exp[-\beta \Delta V]$, at that particular point.

Results

The first question to be answered about the aMD^b approach is if this method is able to improve conformational transitions. To address this question, we performed MD simulations of a butane molecule in explicit water and monitored the dihedral angle shown in Figure 3. Figure 4 shows the result obtained from normal MD simulations. It is worth noting that even for a simple system like this, the number of conformational transitions is still very limited. Figure 4-middle displays the results obtained from the aMD^b approach. In this simulation, only the torsional term of the potential energy was applied in the boost potential (aMDt^b). Parameters E and α were set to 0.5 and 0.2 kcal/mol, respectively. For comparison, Figure 4-bottom shows the dihedral transitions calculated with the aMDt approach. In this case, parameters E and α were set to 5.0 and 0.5 kcal/mol, respectively. As expected, more conformational transitions are observed with aMDt^b and aMDt than with normal MD simulations. Figure 4 also reveals that, even though both methods improved conformational sampling, aMDt^b still produces a much larger number of transitions than aMDt.

Free Energy Calculations. The propane-to-propane system (Figure 5) was used to test convergence and accuracy of the accelerated TI simulations. A similar system, ethane-

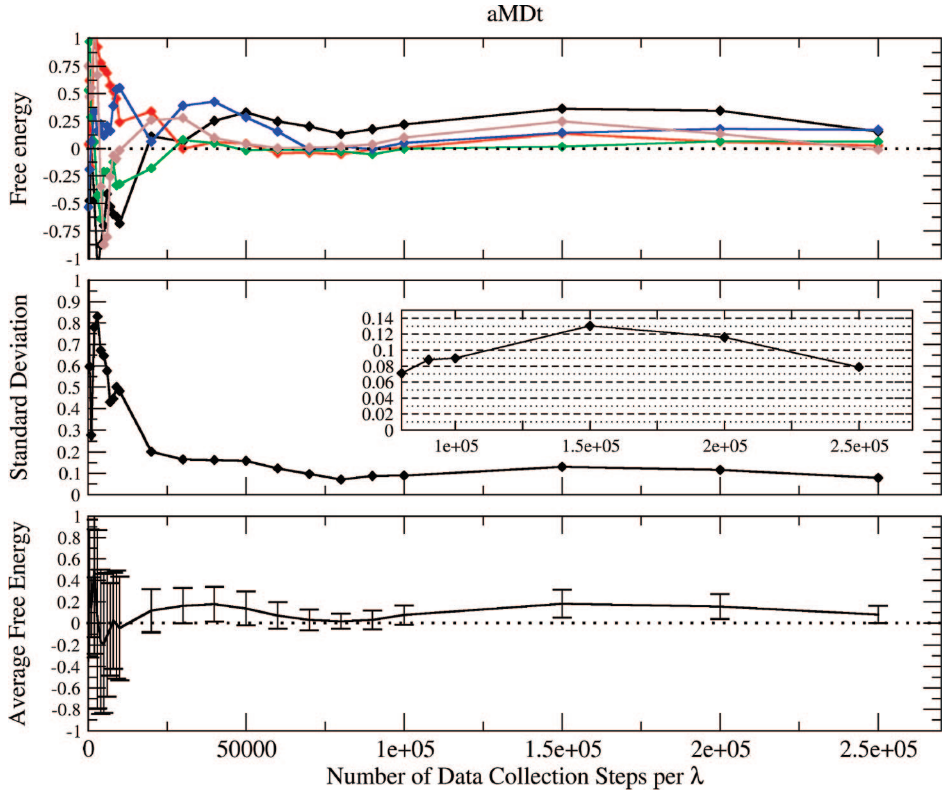


Figure 8. Free energy change in kcal/mol, calculated for the propane-to-propane simulations as a function of time from five independent simulations (top). Standard deviation of the results from the five independent simulations (middle). The inset plot shows in detail the standard deviation as a function of time for the points with at least 80 × 10^3 of data collection steps per λ. The same number of equilibration and data collection steps were used for each λ. Average free energy change and the error associated with each point were calculated from the five independent simulations (bottom). Units are in kcal/mol.

to-ethane transformation, has been used before by other groups to test the performance of different approaches to calculate free energy changes. This transformation is particularly interesting because independent of the force field, water model, or simulation method used, the free energy change should be equal to zero ($\Delta G = 0$).

The free energy changes obtained with normal MD simulations are compared to the ones obtained with aMDt^b, aMDt, and aMDTt^b. Figures 6, 7, 8, 9, and 10 show the free energy change, the average free energy, and the error associated with the propane-to-propane transformation calculated from five independent TI simulations. In all simulations, the same amount of sampling was performed at each window, and an equal amount of time was spent in equilibration and data collection. The error was estimated by calculating the standard deviation of the five independent simulations as a function of time. All TI calculations with normal MD fail to reproduce the expected free energy value (Figure 6), converging to free energy values of ≈−0.4 kcal/mol. The change with time of the average free energy toward the expected free energy value is rather slow, and it is clear that this normal MD requires much longer simulations to reproduce accurate results. Figure 7 displays the TI results obtained with aMDt^b. Although the calculated average free energy change is closer to the corrected free energy value, like normal MD, longer simulations are still required to reproduce the correct average free energy change. Similar results were obtained when the aMDt approach was applied (Figure 8). However aMDt^b still seems to converge better than aMDt. Nevertheless, both approaches perform better than normal TI simulations, and this improvement can be mainly attributed to the increasing in conformational sampling.

As mentioned before, aMDt^b and aMDt approaches modify the energy landscape by adding a boost potential to the potential surface, and, in these cases, the boost potential is based on the torsional terms of the potential energy. Even though the conformational sampling is clearly enhanced in both approaches, those approaches still fail to generate accurate results. The reason for that might be the absence of energy terms in the boost potential describing solute−solvent and solvent−solvent interactions. Therefore, in order to also accelerate the solvent response along the propane-to-propane transformation, the dual boost approach was also tested with TI calculations (ΔV_T was applied with parameters E and α set to −3.0 and 30.0 kcal/mol per atom, respectively). Owing to instabilities introduced by the application of aMDTt approach in TI simulations, only results obtained with aMDTt^b are displayed in Figure 9. By comparing aMDTt^b and aMDt^b, we see clearly that inclusion of the potential energy terms describing solute−solvent and solvent−solvent interactions in eq 3 dramatically improves the accuracy and convergence of the TI simulations. It is worth mentioning that all calculated free energy values using aMDTt^b, with at least 100 ps of data collection, converged to the correct value and are within the estimated error. For the system studied in this work, aMDTt^b was the only approach to achieve

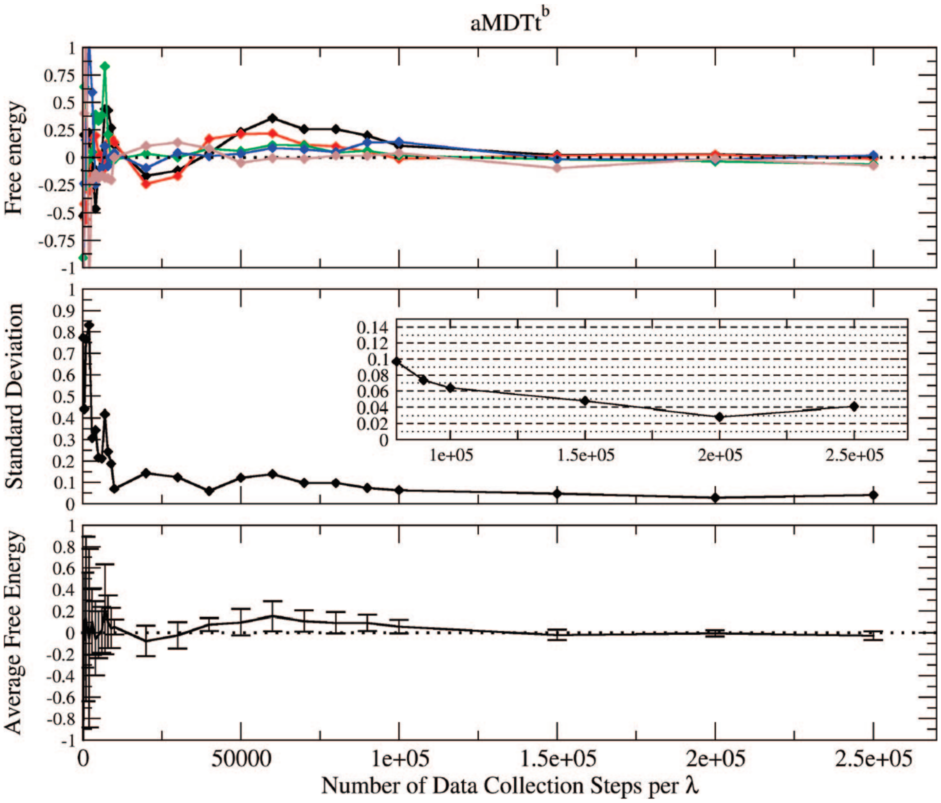


Figure 9. Free energy change in kcal/mol, calculated for the propane-to-propane simulations as a function of time from five independent simulations (top). Standard deviation of the results from the five independent simulations (middle). The inset plot shows in detail the standard deviation as a function of time for the points with at least 80 × 10^3 of data collection steps per λ. The same number of equilibration and data collection steps were used for each λ. Average free energy change and the error associated with each point were calculated from the five independent simulations (bottom). Units are in kcal/mol.

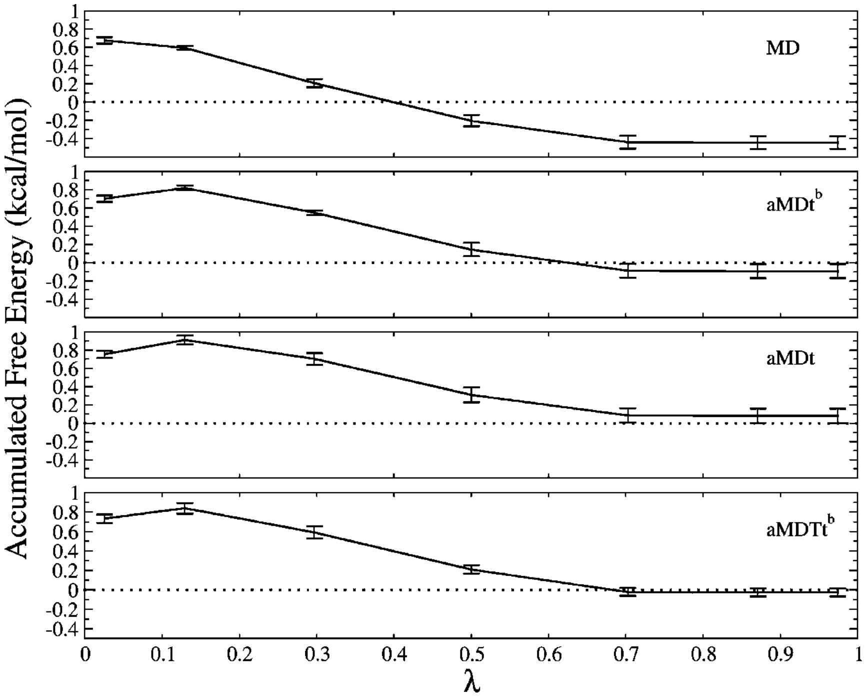


Figure 10. Accumulated free energy change for the propane-to-propane simulations calculated with 250 × 10^3 of data collection steps per λ. The same number of equilibration and data collection steps were used for each λ.

converged and accurate results from TI simulation ($\Delta G = -0.027 \pm 0.04$ kcal/mol).

Figure 10 shows the accumulated free energy for each value of λ. In this plot, the accumulated free energy was calculated by using equilibration = data collection time = 250 ps for each window. Except for the aMDTt^b, which performed remarkably well, all approaches failed to reproduce the corrected free energy change for the propane-to-propane transformation.

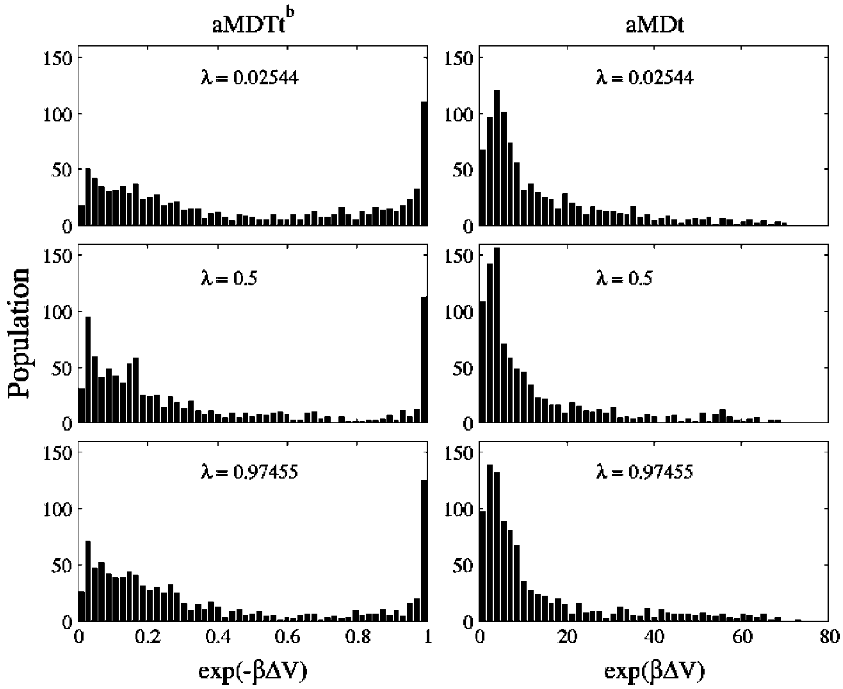


Figure 11. Distribution of Boltzmann factor of the boost potential calculated from propane-to-propane simulations using the aMDTt^b (left) and aMDt (right) approaches.

Distribution of the Boost Energy along the TI Simulations. The main difference between the two approaches (eqs 1 and 3) consists of how the modified potential surface is generated. For methods based on the aMD approach, the molecular motions are accelerated by raising the energy basins on the potential surface. Although this method proved to be excellent to enhance conformational sampling of biomolecules, some issues concerning the calculation of thermodynamics properties still need to be addressed. For instance, to fully recover ensemble average properties, each point in the phase space should be multiplied by its respective Boltzmann factor of the boost energy, exp($\beta\Delta V$). In some cases, when this procedure is applied, relatively few configurations in the entire trajectory have significant contributions to the ensemble average. As a consequence, the statistics are compromised, and the thermodynamic property is not fully converged. This is the main issue to be addressed when the aMD method is coupled with TI calculations. It is worth mentioning that in this implementation ΔV is a nonnegative number, and its respective Boltzmann factor produces numbers in the interval [1→∞). Thus, large values of exp($\beta\Delta V$) correspond to configurations near energy minima of the potential surface, while small values correspond to relatively high-energy regions. Figure 11, on the right, displays the distribution of the boost factor along the aMDt simulations at three different values of λ. It is clear from those plots that, even for this rather low acceleration condition, the system spends almost the entire simulation in regions of relatively high energy. Only very few configurations (for instance configurations with exp($\beta\Delta V$)>50) will effectively contribute to the ensemble average.

In order to address this issue, here, we introduce the aMD^b approach aiming to improve sampling without compromising the statistics in the TI calculations. As mentioned before, in this approach, regions near the minima in the potential surface are left unchanged, and the Boltzmann factor of the boost energy is now defined as exp($-\beta\Delta V$). In this implementation, exp($-\beta\Delta V$) assumes values in the interval (0←1]. Thus, unlike the aMD approach, all configurations near the low-energy regions of the potential surface ($\Delta V = 0$), which are the ones that most contribute to the ensemble average, have the same weight of exp($-\beta\Delta V$) $\approx$ 1. Besides that, configurations sampled in high-energy regions of the conformational space have rather small weights, exp($-\beta\Delta V$) $\ll$ 1, and, as a consequence, have a fairly small contribution to the ensemble average. Figure 11, on the left, shows the boost factor distribution along the aMDTt^b simulations at three different values of λ. It is clear that the aMD^b approach is not only able to sample both low- and high-energy regions of the potential surface but also to keep regions near the minima well populated. As a consequence, the statistics are not compromised in the TI calculations, and the ensemble average is recovered (Figure 9).

Methods

All calculations were performed using the Sander module in the AMBER8[47] package that was modified to carry out the accelerated MD simulations. The GAFF force field was used to describe the solute in all simulations. The butane molecule was solvated in a periodic box of explicit TIP3P waters,[48] which extends on each side 10 Å from the closest atom of the solute, by using the Leap module in AMBER. To bring the system to its correct density, we carried out an MD simulation for 1 ns in which the NPT ensemble (T = 300 K, P = 1 atm) was applied. All data collection was carried out over MD simulations of 1 ns, during which the NVT ensemble (T = 300 K, density= 0.984 g/mL) was applied. The final configuration was then used as the starting point for the propane → propane simulations. In both systems, butane and propane → propane simulations, each solute atom was assigned with zero partial charge. The free energy change was calculated by varying λ form 0 (initial state) to 1 (final state). All TI simulations were carried out using seven discrete points of λ, which were determined by Gaussian quadrature formulas. Normal and accelerated MD simulations of 500 ps were carried out for each λ point. The NVT ensemble was used in all TI simulations. Temperature and pressure were controlled via a weak coupling to external temperature and pressure baths[49] with coupling constants of 0.5 and 1.0 ps, respectively. Apart from all TI simulations where the time step was set to 1 fs, the equations of motion were integrated with a step length of 2.0 fs using the Verlet Leapfrog algorithm.[50] For further analysis, the trajectory was saved every 1.0 ps. The PME summation method was used to treat the long-range electrostatic interactions in the minimization and simulation steps.[51,52] The short-range nonbonded interactions were truncated using a 8 Å cutoff, and the nonbonded pair list was updated every 20 steps.

Conclusions

In this work, we showed a straightforward way of coupling the Thermodynamic Integration approach with the accelerated MD method. We also introduced a new approach, aMD^b, aiming to improve convergence and efficiency of free energy calculations in condensed-phase systems. The results obtained with aMD and aMD^b were compared with conventional TI calculations. Our results showed that both accelerated MD approaches improve conformation sampling when compared to normal MD simulations. When applied to just torsion terms of potential energy, both approaches, aMDt and aMDt^b, increased substantially the number of conformation transitions of the butane molecule in explicit water when compared to normal MD simulations. In addition, the accuracy of free energy simulations was significantly improved when sampling of internal degrees of freedom of solute was enhanced. However, accurate and converged results were only achieved when the solvent interactions were taken into account in the accelerated MD approaches. When combined with aMD^b, the application of dual-boost approach improved markedly the convergence and accuracy of TI calculations. By analyzing the distribution of the boost potential along the free energy simulations, we observed that the aMD^b approach efficiently samples both low- and high-energy regions of the potential surface. Since this approach also maintains well populated regions near the minima, the statistics are not compromised in the TI calculations, and, as a result, the ensemble average can be recovered.

Acknowledgment. This work was supported in part by grants from NSF, NIH, the Center for Theoretical Biological Physics, the National Biomedical Computation Resource, NSF Supercomputing Centers, and Accelrys, Inc.

References

(1) Tembe, B. L.; Mccammon, J. A. *Comput. Chem.* **1984**, *8*, 281.

(2) Gilson, M. K.; Zhou, H. X. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 21.

(3) Straatsma, T. P.; McCammon, J. A. *J. Chem. Phys.* **1991**, *95*, 1175.

(4) VanGunsteren, W. F.; Berendsen, H. J. C. *Angew. Chem., Int. Ed. Engl.* **1990**, *29*, 992.

(5) Jorgensen, W. L. *Acc. Chem. Res.* **1989**, *22*, 184.

(6) Beveridge, D. L.; Dicapua, F. M. *Annu. Rev. Biophys. Biophys. Chem.* **1989**, *18*, 431.

(7) Mezei, M.; Beveridge, D. L. *Methods Enzymol.* **1986**, *127*, 21.

(8) Jorgensen, W. L.; Ravimohan, C. *J. Chem. Phys.* **1985**, *83*, 3050.

(9) Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22*, 1420.

(10) Kirkwood, J. G. *J. Chem. Phys.* **1935**, *3*, 300.

(11) Simonson, T.; Archontis, G.; Karplus, M. *Acc. Chem. Res.* **2002**, *35*, 430.

(12) Kollman, P. *Chem. Rev.* **1993**, *93*, 2395.

(13) Straatsma, T. P.; McCammon, J. A. *Annu. Rev. Phys. Chem.* **1992**, *43*, 407.

(14) Straatsma, T. P.; McCammon, J. A. *Methods Enzymol.* **1991**, *202*, 497.

(15) Adcock, S. A.; McCammon, J. A. *Chem. Rev.* **2006**, *106*, 1589.

(16) Jorgensen, W. L.; TiradoRives, J. *J. Phys. Chem.* **1996**, *100*, 14508.

(17) Min, D. H.; Li, H. Z.; Li, G. H.; Bitetti-Putzer, R.; Yang, W. *J. Chem. Phys.* **2007**, *126*, 144109.

(18) Frantz, D. D.; Freeman, D. L.; Doll, J. D. *J. Chem. Phys.* **1990**, *93*, 2769.

(19) Torrie, G. M.; Valleau, J. P. *J. Comput. Phys.* **1977**, *23*, 187.

(20) Bartels, C.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 865.

(21) Sugita, Y.; Kitao, A.; Okamoto, Y. *J. Chem. Phys.* **2000**, *113*, 6042.

(22) Itoh, S. G.; Okamoto, Y. *J. Chem. Phys.* **2006**, *124*, 104103.

(23) Bussi, G.; Laio, A.; Parrinello, M. *Phys. Rev. Lett.* **2006**, *96*, 090601.

(24) Grubmuller, H. *Phys. Rev. E* **1995**, *52*, 2893.

(25) Lee, J.; Scheraga, H. A.; Rackovsky, S. *J. Comput. Chem.* **1997**, *18*, 1222.

(26) Piela, L.; Kostrowicki, J.; Scheraga, H. A. *J. Phys. Chem.* **1989**, *93*, 3339.

(27) Darve, E.; Wilson, M. A.; Pohorille, A. *Mol. Simul.* **2002**, *28*, 113.

(28) Mitsutake, A.; Sugita, Y.; Okamoto, Y. *Biopolymers* **2001**, *60*, 96.

(29) Kim, J.; Straub, J. E.; Keyes, T. *J. Chem. Phys.* **2007**, *126*, 135101.

(30) Li, H.; Min, D.; Liu, Y.; Yang, W. *J. Chem. Phys.* **2007**, *127*, 094101.

(31) Ceotto, M.; Ayton, G. S.; Voth, G. A. *J. Chem. Theory Comput.* **2008**, *4*, 560.

(32) Xing, C. Y.; Andricioaei, I. *J. Chem. Phys.* **2006**, *124*, 034110.

(33) Hamelberg, D.; Mongan, J.; McCammon, J. A. *J. Chem. Phys.* **2004**, *120*, 11919.

(34) Voter, A. F. *Phys. Rev. Lett.* **1997**, *78*, 3908.

(35) Voter, A. F. *J. Chem. Phys.* **1997**, *106*, 4665.

(36) Hamelberg, D.; de Oliveira, C. A. F.; McCammon, J. A. *J. Chem. Phys.* **2007**, *127*, 155102.

(37) Hamelberg, D.; McCammon, J. A. *J. Am. Chem. Soc.* **2005**, *127*, 13778.

(38) Hamelberg, D.; Mongan, J.; McCammon, J. A. *Protein Sci.* **2004**, *13*, 76.

(39) Hamelberg, D.; Shen, T.; McCammon, J. A. *J. Am. Chem. Soc.* **2005**, *127*, 1969.

(40) Markwick, P. R. L.; Bouvignies, G.; Blackledge, M. *J. Am. Chem. Soc.* **2007**, *129*, 4724.

(41) Pearlman, D. A.; Kollman, P. A. *J. Chem. Phys.* **1991**, *94*, 4532.

(42) Pearlman, D. A. *J. Phys. Chem.* **1994**, *98*, 1487.

(43) Hamelberg, D.; Shen, T.; McCammon, J. A. *J. Chem. Phys.* **2005**, *122*, 241103.

(44) Hamelberg, D.; Shen, T. Y.; McCammon, A. *Biophys. J.* **2005**, *88*, 183A.

(45) Berens, P. H.; Mackay, D. H. J.; White, G. M.; Wilson, K. R. *J. Chem. Phys.* **1983**, *79*, 2375.

(46) Mcquarrie, D. A. *Phys. Today* **1965**, *18*, 74.

(47) Case, D. A.; Perlman, D. A.; Caldwell, J. W.; Chetham, T. E., III; 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.; Kollman, P. A. 2002.

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

(49) Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.

(50) Hockney, R. W. *Bull. Am. Phys. Soc.* **1968**, *13*, 1747.

(51) Ding, H. Q.; Karasawa, N.; Goddard, W. A. *J. Chem. Phys.* **1992**, *97*, 4309.

(52) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577.

CT800160Q

JCTC Journal of Chemical Theory and Computation

Origins of Resistance Conferred by the R292K Neuraminidase Mutation via Molecular Dynamics and Free Energy Calculations

Ricky Chachra[†,§] and Robert C. Rizzo[*,†,‡]

Department of Applied Mathematics and Statistics, and the Institute for Chemical Biology and Drug Discovery, Stony Brook University, Stony Brook, New York 11794

Received February 28, 2008

Abstract: Point mutations in the influenza virus enzyme neuraminidase (NA) have been reported that lead to dramatic loss of activity for known NA inhibitors including the FDA approved sialic acid mimics zanamivir and oseltamivir. A more complete understanding of the molecular basis for such resistance is a critical component toward development of improved next-generation drugs. In this study, we have used explicit solvent all-atom molecular dynamics simulations, free energy calculations (MM-GBSA), and residue-based decomposition to model binding of four ligands with NA from influenza virus subtype N9. The goal is to elucidate which structural and energetic properties change as a result of a mutation at position R292K. Computed binding free energies show strong correlation with experiment ($r^2 = 0.76$), and an examination of individual energy components reveal that changes in intermolecular Coulombic terms (ΔE_{coul}) best describe the variation in affinity with structure ($r^2 = 0.93$). H-bond populations also parallel the experimental ordering ($r = -0.96$, $r^2 = 0.86$) reinforcing the view that electrostatics modulate binding in this system. Notably, in every case, the simulation results correctly predict that loss of binding occurs as a result of the R292K mutation. Per-residue binding footprints reveal that changes in $\Delta\Delta E_{\mathrm{coul}}$ for R292K-wildtype at position 292 parallel the change in experimental fold resistance energies ($\Delta\Delta G_{\mathrm{R292K\text{-}WT}}$) with S03 < S00 < S02 < S01. The footprints also reveal that the most potent ligands have (1) less reliance on R292 for intrinsic affinity, (2) enhanced binding via residues E119, E227, and E277, and (3) flatter ΔE_{coul} and ΔH-bond profiles. Improved resistance for S03 appears to be a function of the ligand's larger guanidinium group which leads to an increased affinity for wildtype NA while at the same time a reduction in favorable interactions localized to R292. Overall, the computational results significantly enhance experimental observations through quantification of specific interactions which govern molecular recognition along the N9-ligand binding interface.

Introduction

Hemagglutinin (HA) and neuraminidase (NA) are glycoproteins on the surface of the influenza virus and, as integral to the life cycle of the virus, are attractive targets for drug design.[1] The World Health Organization (WHO) classifies the HAs into sixteen subtypes (H1−H16) and the NAs into nine subtypes (N1−N9) based on antigenic and genetic analysis.[1–3] Current flu vaccines are based on common circulating influenza A virus subtypes H1N1 and H3N2 and influenza B virus.[4] Overall, seasonal influenza causes an estimated 250,000−500,000 deaths worldwide and about 30,000−50,000 deaths in the United States each year.[4] Historically, prior pandemics include the 1918 Spanish flu (H1N1) with an estimated 50−100 million deaths,[5] the 1957

[*] Corresponding author e-mail: rizzorc@gmail.com.

[†] Department of Applied Mathematics and Statistics.

[‡] Institute for Chemical Biology and Drug Discovery.

[§] Current address: Weill Cornell Graduate School of Medical Sciences, Cornell University, New York, NY 10021.

10.1021/ct800068v CCC: $40.75 © 2008 American Chemical Society
Published on Web 08/12/2008

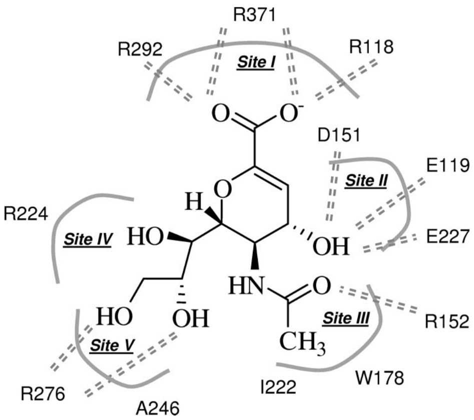
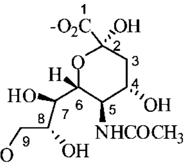
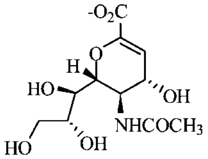
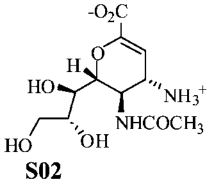


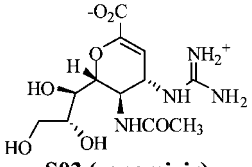
Figure 1. Key interactions in the neuraminidase active sites for inhibitor Neu5Ac2en (Table 1, S01). Figure adapted from ref 25 which divides the site into five regions (Site I−V). Specific hydrogen bonds and salt-bridge interactions are shown as dashed lines.

Table 1. Chemical Structures and Inhibition Constants for Sialic Acid Analogs[a] with Wildtype (WT) and Mutant (R292K) Neuraminidase


S00 (sialic acid)


S01


S02


S03 (zanamivir)

code	K$_i$-WT (μM)[b]	ca. ΔG_{WT} (kcal/mol)[c]	K$_i$-R292K (μM)[b]	ca. ΔG_{R292K} (kcal/mol)[c]	$\Delta\Delta G_{R292K\text{-}WT}$ (kcal/mol)
S00	55	−5.81	1820	−3.74	2.07
S01	2.64	−7.61	280	−4.85	2.76
S02	0.148	−9.32	14	−6.62	2.70
S03	0.002	−11.87	0.033	−10.21	1.66

[a] S00, *N*-acetylneuraminic acid (Neu5Ac, sialic acid); S01, 2-deoxy-2,3-dehydro-*N*-acetylneuraminic acid (Neu5Ac2en, DANA); S02, 4-amino Neu5Ac2en; S03, 4-guanidino Neu5Ac2en (zanamivir). [b] Experimental values from ref 26 for A/NWS/Tern/Australia/G70C (subtype N9). [c] Experimental free energies of binding (ΔG_{WT}, ΔG_{R292K}) estimated as ΔG_{bind} exptl ≈ RT ln (K$_i$ in molar) at 25 °C.

Asian flu (H2N2) with >1 million deaths, and the 1968 Hong Kong flu (H3N2) with ca. 700,000 deaths.[6] The more recent but highly pathogenic avian influenza A subtype H5N1, first isolated in 1998,[7] has an astounding mortality rate with 383 human cases and 241 deaths reported to WHO for the period 2003−June 1, 2008.[3] While HA is involved in binding of the virion to the host cell, NA cleaves terminal sialic acid groups from host cell-surface glycoproteins and glycolipids resulting in release of viral progeny and further spread of infection. A significant focus of drug design against influenza has involved development of derivatives of sialic acid designed to inhibit the release of viral progeny by binding to NA. Notably, numerous computational studies have aided these efforts,[8–20] and development of "flu" inhibitors is often cited as a seminal example of structure-based drug design.[21] To date, two NA inhibitors have been approved by the FDA,[22] oseltamivir and zanamivir, and a third compound, peramivir, is in phase 2 of clinical trials.[23]

In its biologically functional form, NA is a tetramer made of four identical subunits each of which contains a super-barrel structure primarily made of beta-sheets.[24] Each monomer contains an active site which Stoll et al.[25] have divided into five regions (Sites I−V) as shown in Figure 1. Overall the binding site is structurally well-conserved across subtypes and strains and is highly charged; nine of the twelve binding pocket residues in Figure 1 are charged. Of particular importance is the trio of Arg residues in Site I at positions 118, 292, and 371 which form strong hydrogen bonds with the carboxylate off of position C2 on the central ring of ligand substrates (see Table 1 for numbering).[8] It is important to note that this carboxylate group has been a key feature of all reported inhibitors of neuraminidase. The acetamido group (position C5) which interacts with Site III residues is also largely conserved in most inhibitors.

Although the active site of NA is largely conserved across all subtypes (N1−N9) and strains,[8] evolving point mutations in NA pose a major challenge for development of antivirals as several mutations are known to cause a serious loss of sensitivity to reported chemotherapeutics.[26–32] For instance, different strains with N1 subtypes bearing H274Y and N294S mutations have been shown to confer up to 1800- and 200-fold resistance respectively to oseltamivir.[28,29] Point mutations known to adversely affect binding and activity for inhibitors with subtype N9 include R292K (Site I)[26] and E119G (Site II)[27] and to a lesser extent R152K (Site III).[30] Understanding the molecular basis for resistance caused by such deleterious mutations is critical for the development of more effective anti-influenza virus compounds.

McKimm-Breschkin et al.[26] have reported activities for a series of sialic acid mimics which inhibit NA from subtype N9 from influenza strain A/NWS/Tern/Australia/G70C for both the wildtype and R292K mutant. Table 1 shows structures, activities, free energies of binding (ΔG_{WT}, ΔG_{R292K}), fold resistance energies ($\Delta\Delta G_{R292K\text{-}WT}$), and code numbers for S00, *N*-acetylneuraminic acid (Neu5Ac, sialic acid); S01, 2-deoxy-2,3-dehydro-*N*-acetylneuraminic acid (Neu5Ac2en, DANA); S02, 4-amino-Neu5Ac2en; and S03, 4-guanidino-Neu5Ac2en (zanamivir, RELENZA[22]). S00 and the other three ligands differ at position C2 where S00 has a hydroxyl group which results in a nonplanar six-membered ring compared with the other compounds which contain double bonded character at position C2=C3 resulting in a more planar scaffold. Otherwise, the ligands differ only in functionality at position C4. S02 and S03 bear positively charged amino and guanidino functionality, respectively, while S00 and S01 contain a neutral OH group. The presence of a charged group at C4 has a significant effect on interaction of those ligands with residues in the binding site. The ligands in Table 1 are arranged in order of increasing activity. All ligands show a substantial reduction in experimental activities due to the R292K mutation.[26] Although S03 (zanamivir) is the most potent, and the most resilient to R292K, the experimental fold resistance ($\Delta\Delta G_{\text{R292K-WT}}$) energies reveal that the weakest binder S00 is actually the second-most robust to the point mutation (Table 1, S03 > S00 > S02 > S01).

Characterizing binding for ligands with wildtype and mutant forms of NA will ultimately enable the design of improved inhibitors. Prior NA computational studies include rational design[8] using Goodford's GRID program,[33] energy minimization and molecular dynamics (MD),[9,11] Poisson-Boltzmann (PB) calculations,[10] linear interaction energy (LIE) calculations,[12] PMF-scoring using DOCK,[13] comparative binding energy analysis,[14] molecular orbital calculations,[15] MM-PBSA simulations,[16–18] QSAR analysis,[19] charge optimization,[34] and MD simulations aimed at characterizing loop flexibility[20] from recently crystallized[35] N1 subtypes. The present study is focused on characterization of the R292K variant. All-atom explicit solvent MD simulations, free energy calculations, and residue-based decomposition were used to model ligands in complex with NA subtype N9 with the following goals: (1) develop a robust computational model for prediction of binding affinities in agreement with experiment, (2) determine which factors contribute most to the observed binding affinities, and (3) delineate which specific structural and energetic factors contribute to the R292K resistance profiles. Well-tested computational models of inhibitors with N9, and clinically relevant mutants, will enable isolation of the energetic and structural determinates which confer improved binding resilience of S03 and a greater understanding of what drives molecular recognition with NA in general. Development of improved inhibitors across all NA subtypes, including the recently discovered highly pathogenic avian strain,[7] is paramount given the likelihood of future influenza pandemics.[6,36,37]

Theoretical Methods

In this study, free energies of binding were estimated for four ligands with wildtype neuraminidase and an R292K mutant using the single trajectory Molecular Mechanics Generalized Born Surface Area (MM-GBSA) method.[38,39] This approach was recently used to successfully investigate binding for a series of large viral entry peptide inhibitors of HIVgp41[40] and to determine the origins of selectivity for small inhibitors of matrix metalloproteases.[41] Although considered an approximate free energy calculation technique, the benefits include relative ease of setup and use, the ability to study large structural changes, and the ability to compare binding energies between ligands with diverse topologies. Tradeoffs include an incomplete accounting of all solute entropic effects and the fact that an implicit solvent model is used for the free energy calculations. However, changes in solute configurational entropies can be reasonably assumed to remain constant when ligands have similar binding poses and these changes are often ignored. Further, recent advances[42,43] and robust evaluation[44] of GBSA continuum methods for estimation of desolvation effects have revealed that implicit solvent models can indeed be very accurate provided that correct charge models and radii are employed in the calculations. Finally, the single trajectory method used here obviates the need for alchemical transformations to obtain free energies given that only a single simulation of each protein−ligand complex is required.[38,39] Thus, results can be obtained considerably faster than what are historically regarded as gold-standard free energy calculation methods which rely on perturbation techniques such as thermodynamic integration (TI) and free energy perturbation (FEP).[45,46] Despite the approximation made in MM-GBSA, the method has been used with good overall success to study a wide variety of problems.[18,40,41,47–50] The present study serves as an additional test of the utility of the method for estimation of the effects of point mutations on protein−ligand binding.

The calculations make use of explicit solvent MD simulations to generate ensembles of low energy structures which are postprocessed to compute the binding energy components. Implicit solvent is used only for estimation of the desolvation terms. After the simulations, explicit solvent is stripped off, and the coordinates for each species are separated to yield the complex, unbound receptor, and unbound ligand. Average energies (and associated uncertainties) are computed from many single point calculations using the ensemble of structures saved periodically during the MD simulations with each species total free energy estimated using eq 1. The total binding free energy is computed using eq 2.

$$G = \Delta G_{\text{hyd}} + E_{\text{MM}} - \text{TS} \tag{1}$$

$$\Delta G_{\text{bind}} = G_{\text{complex}} - (G_{\text{receptor}} + G_{\text{ligand}}) \tag{2}$$

Free energy of hydration ($\Delta G_{\text{hyd}} = G_{\text{polar}} + G_{\text{nonpolar}}$) terms which account for the desolvation penalties which occur upon binding are estimated from Generalized Born (GB) and Solvent Accessible Surface Area (SASA) calculations which yield G_{polar} and G_{nonpolar}, respectively. As validation, Rizzo et al.[44] have recently shown good agreement between experiment and theory from calculations of more than 500 organic molecules using continuum GBSA methods to estimate ΔG_{hyd}. The E_{MM} term represents the sum of electrostatic (Coulombic), van der Waals (Lennard-Jones), and internal energies (bonds, angles, and dihedrals) which are computed using the same molecular mechanics force field used during the original MD simulations. Using the single trajectory approximation, coordinates of separated ligand and receptor are identical to those in the complex, thus any changes in bond, angle, and dihedral (ΔE_{bond}, ΔE_{angle}, and $\Delta E_{\text{dihedral}}$) energies inherent in eqs 1 and 2 will cancel and ΔE_{coul} and ΔE_{vdw} reflect only the nonbonded intermolecular energies. The final *TS* terms, representing temperature (*T*) and solute entropy (*S*), were omitted given the consistent binding pose and size for the four NA ligands being studied. Neglecting $T\Delta S$ and internal strain energy is considered to be a reasonable approximation when ligands are of similar size and structure and only relative binding affinities are of interest. However, it should be emphasized that changes in

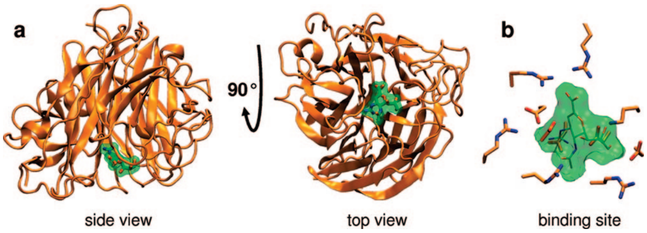


Figure 2. **(a)** Two views of neuraminidase (orange) complexed with ligand S03 (green). Protein coordinates from PDB entry 1F8B. **(b)** Close up view of the highly flexible and charged binding site.

system entropy due to the hydrophobic effect are in fact included as they are inherently contained in the ΔG_{hyd} terms.

Computational Details

System Setup. The biological form of NA is a tetramer; however, each monomer contains a functionally complete binding site,[1] thus only a single monomer was used for the simulations (Figure 2). The binding site is lined with highly charged and flexible Arg, Lys, Asp, and Glu residues. Receptor and ligand preparation was done using the Molecular Operating Environment (MOE) program.[51] The coordinates of neuraminidase subtype N9 complexed with ligand S01 were taken from the 1.80 Å crystallographic structure reported by Smith et al.[52] (pdb entry 1F8B) shown in Figure 2. The N9 receptor structure is from strain A/NWS/Tern/Australia/G70C which is the same strain employed in the experimental activity measurements reported by McKimm-Breschkin et al.[26] (see Table 1) and for the ligands being studied here. Crystallographic water molecules were deleted from 1F8B; however, a single calcium ion near the active site was retained.

Examination of available crystal structures for the 4 ligands complexed with wildtype NA and R292K shows the same well-defined binding pose, and protein side chain conformations in the NA binding site are relatively consistent across the series with the only primary difference being the rotameric state of Lys for the mutation. Therefore, a single set of receptor coordinates (in this case 1F8B with ligand S01) was used as the basis for construction of all simulations. This was primarily motivated by the fact that this would potentially eliminate noise in the simulations caused by multiple different starting conditions (i.e., multiple crystal structures). The R292K mutant was made by manually mutating Arg to Lys at residue 292 in 1F8B and orienting Lys to mimic the rotamer found in structures of R292K and avoid any steric clashes. Initial geometries for analogs S00, S02, and S03 were obtained from crystallographic complexes 2QWB, 2QWD, and 2QWE and oriented into the 1F8B reference frame (which contained S01) through alignment of C-alpha backbone atoms common to all NA structures. This procedure results in eight receptor−ligand complexes constructed from a single set of N9 coordinates.

The wildtype and R292K receptors were saved as PDB files without hydrogen atoms, and the ligands were saved as MOL2 files which included hydrogen atoms. Simulation ready parameter files were constructed for each system using the AMBER8 suite of programs.[53] The *antechamber* and *tleap* modules were used to assign the GAFF[54] force field parameters to the ligands and FF99SB[55] parameters and hydrogen atoms to the receptors. Protein side chains were assigned default AMBER protonation states (Asp/Glu minus, Arg/Lys ±) which, in the current study, yielded good results. A previous study by Masukawa et al.[16] similarly used default protonation states for NA, with a related AMBER force field, also with good success. The current work does not study changes in protonation; however, Smith et al.[52] and Fornabaio et al.[56] have reported computational results for NA under varying states. For the ligands, GAFF parameters were augmented with ChelpG[57] partial atomic charges computed at the HF/6-31G*//HF/6-31G* level of theory using the program Gaussian98.[58] Inspection of energy minimized ligands using the force field in the unbound state revealed a nonplanar guanidino group for S03 which was remedied through manual addition of GAFF improper dihedral angle parameters which forced the group to be planar. Each system was then subjected to a short energy minimization which yielded relatively small changes in geometry (bound and unbound) and no discernible steric clashes which indicated that the starting coordinates and force field parameters were reasonable.

MD Simulations and Postprocessing. Each solvated protein−ligand system contained 390 residues (including one calcium ion and the ligand) and 11,949 TIP3P[59] waters in a rectangular periodic box of 68 × 77 × 77 Å^3. All energy minimizations and MD employed the *sander* module from AMBER8. A nine step equilibration protocol was used prior to production MD in the following order. First, energy minimization for 1000 cycles followed by 50 ps of MD at 298.15 K was performed on each complex using a restraint weight of 5.0 kcal/mol Å^2 on all heavy atoms (steps 1 and 2). This was followed by three rounds of energy minimization for 1000 cycles each in which the restraint weight on heavy atoms was reduced from 2.0, to 0.1, to 0.05 kcal/mol Å^2 (steps 3−5). An additional three rounds of MD (50 ps each at 298.15 K) were performed with decreasing restraint weights reduced from 1.0, to 0.5, to 0.1 kcal/mol Å^2 (steps 6−8). A final equilibration of 50 ps of MD at 298.15 K using restraints only on protein backbone atoms (C-alpha, C, N, O) was then performed using a weight of 0.1 kcal/mol Å^2 (step 9). The production run employed the same weak backbone restraints as the last equilibration step for a total of 2000 ps of MD. A 1 fs time step was used for the equilibration stages (steps 1−9), and a 2 fs was used for the final production runs.

Weak backbone restraints were employed in the final production runs based on preliminary results from unrestrained MD simulations of NA monomers using only implicit solvent which showed larger than expected movement especially for protein termini regions. Larger motion would be expected due to a lack of friction in implicit solvent dynamics but is probably also a consequence of the fact that an NA monomer was simulated instead of a tetramer of NA subunits which would have otherwise held the protein termini restrained. In order to avoid similar artifacts in the explicit solvent TIP3P-MD, a weak restraint was employed to keep the backbone fold intact. Weak restraints have previously been used with good results for simulations of inhibitors with MMPs[41] and HIVgp41.[40] Temperature and pressure of the simulations was regulated using the Berendsen[60] schemes using the heat bath coupling and pressure relaxation time constants of 1.0 ps each. The SHAKE[61] algorithm was applied to constrain bonds involving hydrogen atoms, and the particle mesh Ewald (PME)[62] method was used with 8.0 Å direct-space nonbonded cutoff.

Coordinates (snapshots) of each complex were saved every picosecond (2000 snapshots total) during the MD production trajectory. Each trajectory was then split into separate species representing the complex, the unbound receptor, and the unbound ligand, and the *sander* module was used to perform single point postprocessing calculations to compute the energy components (ΔE_{vdw}, ΔE_{coul}, $\Delta\Delta G_{\mathrm{polar}}$, $\Delta\Delta G_{\mathrm{nonpolar}}$) needed for estimation of the total binding free energy (ΔG_{bind} calcd). The polar energy terms G_{polar} were obtained via the AMBER implementation[63,64] of the Hawkins, Cramer, and Truhlar[65,66] pairwise Generalized Born[67] model as modified by Onufriev et al.[68] (model type igb=5). GB calculations employed dielectric constants of 1 and 78.5 and AMBER mbondi2 radii. Nonpolar terms were estimated as $G_{\mathrm{nonpolar}} = \gamma\mathrm{SASA} + \beta$ with SASA in Å^2 using standard values of $\gamma = 0.00542$ kcal/mol Å^2 and $\beta = 0.92$ kcal/mol.[69] Binding site footprints were obtained from pairwise decomposition of the per-residue interaction energies between the ligands and each NA residue and averaging over the total trajectory. Intermolecular hydrogen bonds were also computed and were defined as a structural interaction between a donor (H_D) and acceptor (X_A) with a distance of 2.5 Å or less and an angle between X_D-H_D---X_A of between 120 and 180°.

Results and Discussion

Simulation Stability. The stability of each simulation was monitored through examination of structural and energetic properties which occurred during the course of the 2 ns production trajectories. Figure 3, which is representative, shows results for ligand S01 complexed with wildtype neuraminidase. Here, plots of root-mean-square-deviation from the original starting coordinates (rmsd, Figure 3a), instantaneous changes in desolvation ($\Delta\Delta G_{\mathrm{hyd}}$, Figure 3b), electrostatic (ΔE_{coul}, Figure 3c), and van der Waals (ΔE_{vdw}, Figure 3d) nonbonded interaction energies are well-behaved which indicate the simulations are reasonably converged. In particular, ΔE_{vdw} interactions remain almost constant across the trajectory, and rmsd values (Figure 3a) show only minor variation. Interestingly, a relatively small shift in ligand positional rmsd (Figure 3a, blue line) at around 300 ps results in a rather large increase in favorable ΔE_{coul} with a concurrent increase in unfavorable $\Delta\Delta G_{\mathrm{hyd}}$. Such a large change in energy, from a relatively small change in geometry, is expected to be a consequence of the fact the NA binding site is so highly charged. After this change, the ligand rmsd remains constant, but the opposing desolvation and electrostatic terms then slowly come back to their respective equilibrium positions observed before the change. Another shift in group correlated energies for ΔE_{coul} with $\Delta\Delta G_{\mathrm{hyd}}$ appears starting at around 1700 ps (Figure 3b vs Figure 3c).

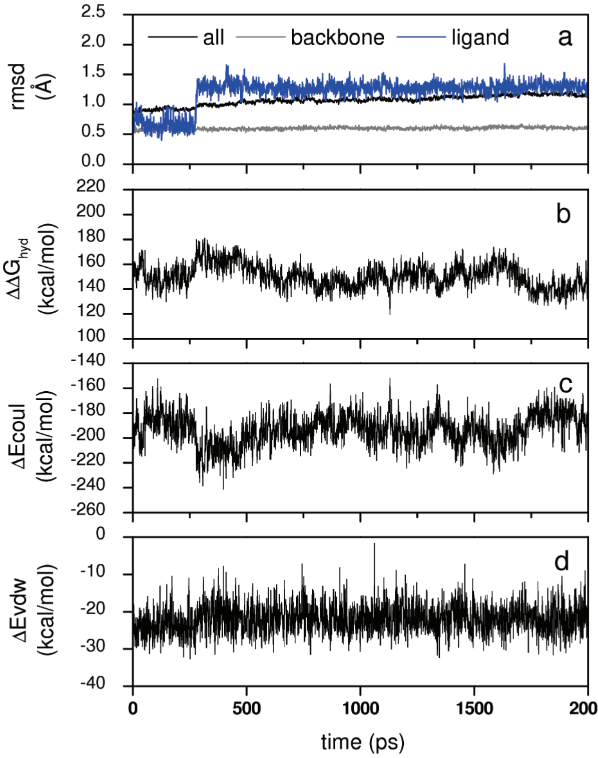


Figure 3. Instantaneous results from MD simulations of ligand S01 with wildtype neuraminidase subtype N9 plotted vs time. **(a)** shows root-mean-square deviation (rmsd) in angstroms (Å) between snapshots from the MD simulations and the initial starting coordinates for all protein heavy atoms (black line), protein backbone main chain atoms C-alpha, C, N, O (gray line), and ligand heavy atoms (blue line). **(b)** shows the change in free energy of hydration ($\Delta\Delta G_{\mathrm{hyd}} = \Delta\Delta G_{\mathrm{polar}} + \Delta\Delta G_{\mathrm{nonpolar}}$), while panels **(c)** and **(d)** show the nonbonded intermolecular electrostatic (ΔE_{coul}) and van der Waals (ΔE_{vdw}) interaction energies, respectively.

For this second period, normal protein side-chain sampling is expected to be the primary factor given the relatively flat ligand rmsd shown in Figure 3a (blue line). In terms of magnitude, the nature of the highly charged NA binding site results in significantly more favorable ΔE_{coul} interaction energies than ΔE_{vdw} (Figure 3c vs Figure 3d). As discussed below, variation in computed electrostatic properties appear to play the dominant role in describing variation in the experimentally observed activities.

Our group and others have previously noted that favorable intermolecular electrostatic energies are anticorrelated to the opposing desolvation penalties.[11,40,41,70–73] Here, the correlation coefficient is computed to be $r = -0.86$ between $\Delta\Delta G_{\mathrm{hyd}}$ and ΔE_{coul} for the 2000 instantaneous energies shown in Figure 3 for S01 with wildtype N9. The subtle interplay of structure with the various energy terms is a hallmark of the intimate relationship between opposing interactions (i.e., desolvation with electrostatics) which ultimately contribute to the overall free energy of binding. Notably, the present simulations and protocols appear to capture such subtleties well. The dramatic effects of desolvation are even more pronounced when considering all four ligands (S00−S03) with both receptors (wildtype and R292K). Here, an $r^2 =$

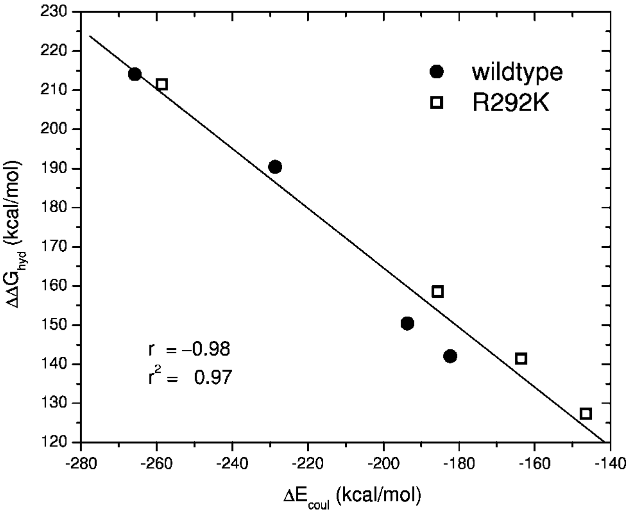


Figure 4. Intermolecular protein−ligand Coulombic energies (ΔE_{coul}) versus opposing desolvation penalties ($\Delta\Delta G_{\text{hyd}}$). Each symbol represents the average energy computed from 2000 MD snapshots saved during simulations of S00−S03 with wildtype (●) and R292K mutant (◘) neuraminidase subtype N9.

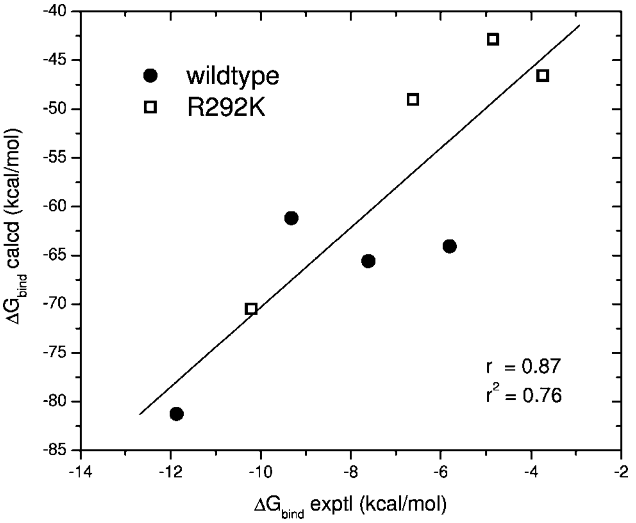


Figure 5. Average computed free energies of binding (ΔG_{bind} calcd) versus experimental activities (ΔG_{bind} exptl) for sialic acid inhibitors with wildtype (●) and R292K mutant (◘) neuraminidase subtype N9.

0.97 is obtained using averaged values ($N = 2000$) for ΔE_{coul} and $\Delta\Delta G_{\text{hyd}}$ as shown in Figure 4.

Correlation with Experimental Activities. Figure 5 shows the correlation between the experimental (ΔG_{bind} exptl) and theoretical (ΔG_{bind} calcd) free energies of binding computed using the MM-GBSA method. Here, each data point represents the average values of ΔG_{bind} calcd obtained from 2000 MD snapshots of the four ligands with either wildtype (filled circles) or the R292K mutant (open squares), and the correlation coefficients of $r = 0.87$ and $r^2 = 0.76$ indicate overall good agreement with the experimental binding free energies. Notably, the calculations correctly predict that the R292K mutations reduced binding with each ligand in every case as highlighted in Figure 6 which shows instantaneous computed ΔG_{bind} values for wildtype (solid lines) and R292K (dashed lines) trajectories. As was observed in plots of individual energetic components (Figure 3), the total free energies (ΔG_{bind} calcd $= \Delta E_{\text{vdw}} + \Delta E_{\text{coul}} + \Delta\Delta G_{\text{hyd}}$) are also well-behaved. Smoothed lines in Figure 6 represent running block averaging over the previous 100 MD snapshots.

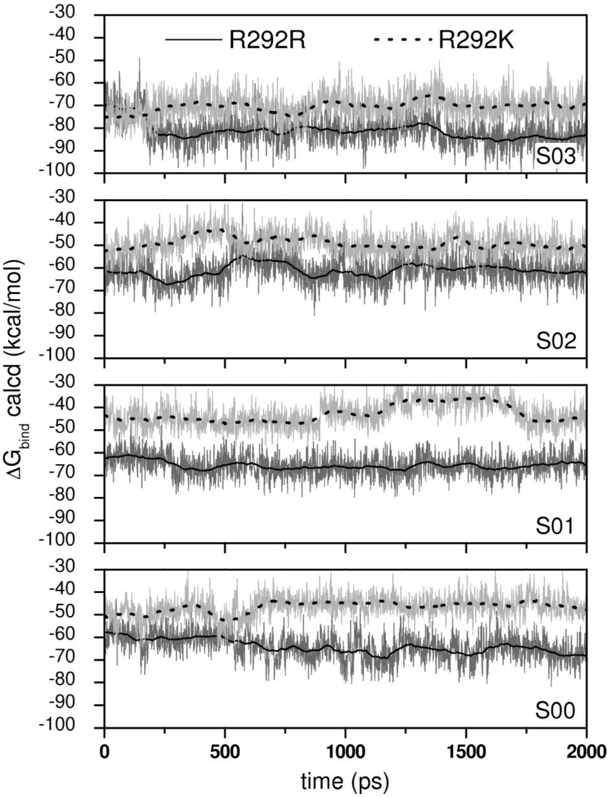


Figure 6. Instantaneous (jagged lines) and 100-block averaged (smoothed lines) free energies of binding (ΔG_{bind} calcd) vs time for ligands with wildtype (black, solid lines) and R292K (gray, dashed lines) neuraminidase subtype N9.

Energy Decomposition. Individual energy terms which contribute to ΔG_{bind} calcd (eqs 1 and 2) were examined to determine which factors drive association and correlate best with the experimental activities and are shown in Table 2. Correlations coefficients (r^2 values) are shown for Coulombic (ΔE_{coul}), van der Waals (ΔE_{vdw}), polar ($\Delta\Delta G_{\text{polar}}$), nonpolar ($\Delta\Delta G_{\text{nonpolar}}$), total electrostatics ($\Delta G_{\text{electro}} = \Delta E_{\text{coul}} + \Delta\Delta G_{\text{polar}}$), and the total computed binding energy (ΔG_{bind} calcd $= \Delta E_{\text{coul}} + \Delta E_{\text{vdw}} + \Delta\Delta G_{\text{polar}} + \Delta\Delta G_{\text{nonpolar}}$).

As previously reported, nonbonded van der Waals interactions often correlate strongly with binding across a variety of sytems.[40,41,74–76] Representing intermolecular packing, favorable ΔE_{vdw} is a good predictor of likely intermolecular geometries. Programs such as DOCK[77] which employ scoring functions where ΔE_{vdw} terms tend to dominate have yielded good success for prediction of binding poses.[78,79] In the present study, the most potent compound S03 is computed to make the strongest ΔE_{vdw} interactions with wildtype neuraminidase compared with the other ligands ($\text{S03}_{\text{wildtype}} = -29.26$ versus ca. -21 to -24 kcal/mol, Table 2, ΔE_{vdw}). But, the trend is not maintained for the R292K mutation (Table 2, $\text{S03}_{\text{R292K}} = -22.92$ versus ca. -21 to -24 kcal/mol), and the overall correlation coefficient for

Table 2. Contributions toward Calculated Free Energies of Binding (ΔG_{bind} calcd) from MD Simulations for Sialic Acid Inhibitors (S00−S03) with Wildtype and R292K Mutant Neuraminidase Subtype N9[a]

	ΔE_{vdw}	ΔE_{coul}	$\Delta\Delta G_{\mathrm{polar}}$	$\Delta\Delta G_{\mathrm{nonpolar}}$	$\Delta G_{\mathrm{electro}}$ =	ΔG_{bind} calcd =	ΔG_{bind} expt ≈
system	A	B	C	D	B+C	A+B+C+D	RT ln(K$_i$)[b]
S00$_{\mathrm{wildtype}}$	−23.62 &±mn; 0.09	−182.32 &±mn; 0.31	146.36 &±mn; 0.23	−4.35 &±mn; 0.002	−35.96	−64.07 &±mn; 0.13	−5.81
S01$_{\mathrm{wildtype}}$	−21.93 &±mn; 0.09	−193.71 &±mn; 0.29	154.70 &±mn; 0.22	−4.28 &±mn; 0.002	−39.01	−65.58 &±mn; 0.11	−7.61
S02$_{\mathrm{wildtype}}$	−22.84 &±mn; 0.11	−228.66 &±mn; 0.33	194.81 &±mn; 0.18	−4.41 &±mn; 0.002	−33.85	−61.18 &±mn; 0.14	−9.32
S03$_{\mathrm{wildtype}}$	−29.26 &±mn; 0.10	−265.78 &±mn; 0.29	218.79 &±mn; 0.18	−4.70 &±mn; 0.001	−46.99	−81.26 &±mn; 0.15	−11.87
S00$_{\mathrm{R292K}}$	−23.98 &±mn; 0.10	−163.63 &±mn; 0.33	145.77 &±mn; 0.27	−4.36 &±mn; 0.003	−17.86	−46.56 &±mn; 0.11	−3.74
S01$_{\mathrm{R292K}}$	−23.31 &±mn; 0.08	−146.52 &±mn; 0.25	131.67 &±mn; 0.19	−4.37 &±mn; 0.001	−14.85	−42.85 &±mn; 0.13	−4.85
S02$_{\mathrm{R292K}}$	−21.40 &±mn; 0.10	−185.63 &±mn; 0.29	162.95 &±mn; 0.21	−4.36 &±mn; 0.002	−22.68	−49.02 &±mn; 0.12	−6.62
S03$_{\mathrm{R292K}}$	−22.92 &±mn; 0.10	−258.67 &±mn; 0.30	216.12 &±mn; 0.20	−4.63 &±mn; 0.001	−42.54	−70.49 &±mn; 0.13	−10.21
	$r^2 = 0.23$	$r^2 = 0.93$	$r^2 = 0.88$	$r^2 = 0.62$	$r^2 = 0.73$	$r^2 = 0.76$	

[a] All energies &±mn; standard error of the mean in kcal/mol computed from each ensemble of 2000 MD snapshots. [b] Activity values from Table 1.

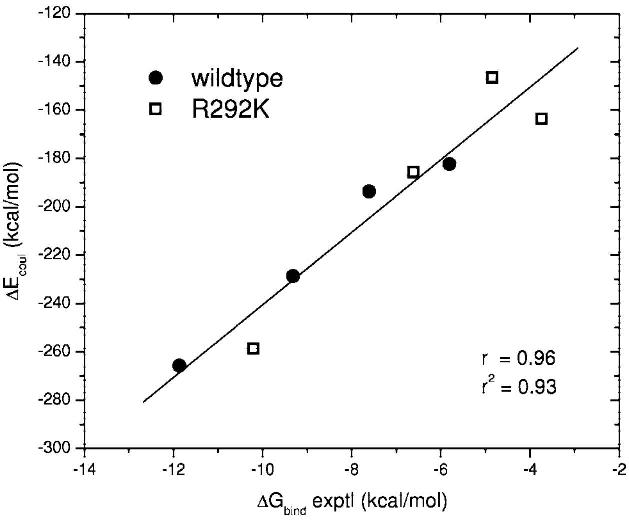


Figure 7. Correlation of nonbonded electrostatic (ΔE_{coul}) interaction energies with experimental activities (ΔG_{bind} exptl) for sialic acid inhibitors with wildtype (●) and R292K mutant (□) neuraminidase subtype N9.

ΔE_{vdw} with experiment is low ($r^2 = 0.23$). However, the ΔE_{coul} terms are computed to be highly correlated with ΔG_{bind} exptl (Table 2) with strong overall correlation coefficients of $r = 0.96$ and $r^2 = 0.93$ (Figure 7). A least-squares fit using ΔE_{coul}, ΔE_{vdw}, $\Delta\Delta G_{\mathrm{polar}}$, and $\Delta\Delta G_{\mathrm{nonpolar}}$ in Table 2 as descriptors for ΔG_{bind} exptl confirms the importance of ΔE_{coul}. To obtain relative weights of the descriptors, each column of raw data in Table 2 was mean centered and scaled by the standard deviation prior to fitting. The large positive coefficient for ΔE_{coul} (1.131) vs the small coefficients obtained for other descriptors ΔE_{vdw} (0.076), $\Delta\Delta G_{\mathrm{polar}}$ (0.094), and $\Delta\Delta G_{\mathrm{nonpolar}}$ (−0.135) clearly indicates Coulombic energy best describes the variation. In addition, the above fit using all four terms yielded a correlation coefficient of $r^2 = 0.93$ with experiment which is the same as that obtained using only the ΔE_{coul} term alone (see Table 2).

Early studies by Taylor and von Itzstein[11] also found that electrostatic terms best described variation in activity for a series of compounds with neuraminidase of subtype N2. The Taylor study obtained $r^2 = 0.80$ using the sum of intermolecular Coulombic and reaction field free energy (computed from Poisson-Boltzmann calculations) which is comparable with the value of $r^2 = 0.73$ obtained here for the total electrostatics term ($\Delta G_{\mathrm{electro}} = \Delta E_{\mathrm{coul}} + \Delta\Delta G_{\mathrm{polar}}$, Table 2). Similarly, Bonnet and Bryce[18] found that total electrostatics ($\Delta G_{\mathrm{electro}}$) yielded the best fit with experiment ($r^2 = 0.72$, called Model 5E) for a series of 10 ligands with wildtype N9 using a MM-GBSA single-step perturbative simulation approach. As noted earlier, desolvation and Coulombic terms are highly anticorrelated (Figure 4), thus variation in $\Delta\Delta G_{\mathrm{polar}}$ is strongly correlated with ΔG_{bind} exptl as expected ($r^2 = 0.88$). The final term, which reflects burial of surface area upon complexation for both ligand and protein ($\Delta\Delta G_{\mathrm{nonpolar}}$), does not correlate as strongly with the activities (overall $r^2 = 0.62$, Table 2). However, $\Delta\Delta G_{\mathrm{nonpolar}}$ is computed to be most favorable for the largest ligand S03 (−4.7 kcal/mol for WT and −4.6 kcal/mol for R292K) compared with the other ligands (ca. −4.3 to −4.4 kcal/mol) which is physically reasonable. Overall, the energetic decomposition suggests that inhibitory activity of sialic acid mimics for wildtype and the R292K mutant is primarily controlled by, and best described by, intermolecular Coulombic interactions (ΔE_{coul}). The negative sum for $\Delta G_{\mathrm{electro}}$, which can be thought of as the intermolecular Coulombic energies mediated by the polar desolvation penalties, is generally larger in magnitude than ΔE_{vdw} (Table 2). This further suggests that electrostatics dominate association.

In contrast, Armstrong et al.[34] found that similar terms (called $\Delta G^{\mathrm{ref}}{}_{\mathrm{elec}}$ in ref 34) were positive for a series of ligands with neuraminidase and concluded that additional factors such as van der Waals, entropy, and molecular strain energies would in many cases be more significant than electrostatics. Strain effects in the NA system have been investigated previously by Masukawa et al.[16] who reported that S00 (sialic acid) could pay ca. +5 kcal/mol penalty (called $\Delta E_{\mathrm{internal}}$ in ref 16) in going from the unbound state to the boat/twist-boat conformation observed in the bound state. Interestingly, their calculations for S01 (DANA) yielded the opposite effect with a net gain of ca. −5 kcal/mol. However, since the experimental binding difference between S00 and S01 is 1.8 kcal/mol (Table 1), which is much smaller than the ~10 kcal/mol difference in $\Delta E_{\mathrm{internal}}$ reported by Masukawa et al.,[16] other energy terms (i.e., ΔE_{coul}, ΔE_{vdw}, $\Delta\Delta G_{\mathrm{hyd}}$) are clearly important. In the present study, the overall good agreement between the simulation results with experiment suggests that accounting for molecular strain is not necessary to accurately rationalize binding in this system. Differences here with studies that suggest other terms are as or more important than ΔE_{coul} could be attributed to changes in number and type of ligands studied, the type of simulations (MD vs single-point), the type of energetic analysis (GB vs PB), or other calculation protocols. The observation that we also obtain strong correlation for $\Delta G_{\text{electro}}$ with ΔG_{bind} exptl (Table 2, $r^2 = 0.73$) reinforces the view that electrostatics both best describe variation and drive association in this system. H-bonding analysis (shown below) provides additional support. Given the highly charged nature of the NA binding site, strong agreement between ΔG_{bind} exptl and ΔE_{coul} (or with solvent mediated $\Delta G_{\text{electro}}$), as opposed to ΔE_{vdw}, is we believe physically reasonable and not unexpected.

Origins of Resistance: Binding Site Footprints. Residue-based decomposition can be used to determine hotspot regions within a binding site and reveal which specific amino acids play important roles for binding. Strockbine and Rizzo[40] recently showed the utility of using such binding site "footprints" to highlight that differential association of peptide inhibitors with HIVgp41 is driven solely by changes in ΔE_{vdw} energies occurring within a highly conserved binding pocket, supporting the hypothesis that the gp41 pocket region is an important drug target site. For the NA system, prior studies have reported a variety of residue-based methods including energy decomposition, average distances, and partial-least-squares analysis in an effort to gain insight into which inhibitor interactions are most important. [9,14,16,17,33] In the present study, Coulombic footprints were analyzed given that variation in ΔE_{coul}, as opposed to ΔE_{vdw} terms, correlates most strongly with the experimental activities ($r^2 = 0.93$, Table 2). H-bonding footprints were also computed. The primary focus here is to delineate the origins of the different fold resistance profiles for the four ligands with R292K. For the ligands with neutral functionality at position C4 (see Table 1 for numbering), Figure 8 highlights that very similar Coulombic footprints are obtained for S00 and S01, and these are both distinctly different from the corresponding footprints for ligands S02 and S03. The more rugged footprints for S00 and S01 show that many receptor residues on NA interact unfavorably (positive ΔE_{coul}) with these ligands as opposed to S02 and S03 which have much smoother ΔE_{coul} profiles and fewer per-residue energies which are unfavorable. This observation is a consequence of the fact that positively charged amine and guanidino groups on S02 and S03 lead to an overall net formal ligand charge of zero as opposed to a net negative charge for S00 and S01.

Focusing in on the key residues in the NA binding site, Figure 9 shows reduced binding footprints, defined here as the subset of residues which have a significant per-residue Coulombic energy contribution (favorable or unfavorable with $\Delta E_{\text{coul}} \geq 20$ kcal/mol). Figure 9a dramatically highlights the similar energetic profiles (ΔE_{coul}) for binding of the negatively charged S00 and S01 with wildtype NA (solid lines) versus neutral ligands S02 and S03 (dashed lines). The Figure 9a footprint quantifies the importance of the Site I central Arg (see Figure 1) at position 371 which is computed here to be the most energetically favorable interaction for each inhibitor. The relative importance of other Site I residues

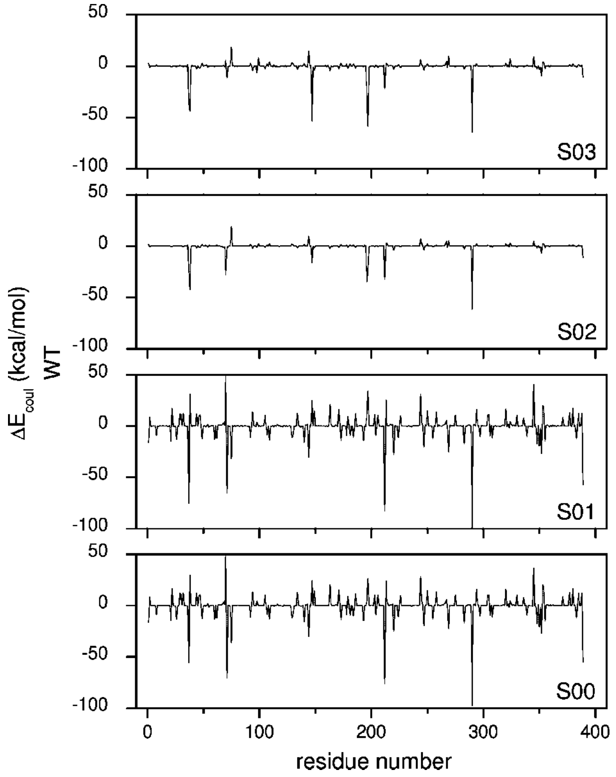


Figure 8. Intermolecular ligand–protein (per-residue) Coulombic energy footprints for wildtype N9.

R118 and R292 which flank R371 are more varied. Figure 9a also reveals that residue R292 in wildtype NA makes strong interactions with ligands S00 and S01 (ca. 80 kcal/mol), but for S02 and S03 (ca. 20–30 kcal/mol) the interaction is less significant (black arrow). Despite binding more tightly with wildtype NA (Table 1), ligands S02 and S03 do not appear to rely as strongly on electrostatic interactions with residue R292 as do S00 and S01 for their intrinsic binding affinity. Thus, the calculations suggest that any loss of interactions at position 292 would not be as detrimental for S02 and in particular S03. In general agreement with this hypothesis, experimentally, S03 loses the least amount of binding energy as a result of the R292K mutation with fold resistance $\Delta\Delta G_{\text{R292K-WT}} = 1.66$ kcal/mol compared with S02 = 2.70 kcal/mol and S01 = 2.76 kcal/mol (Table 1). Interestingly, the experimental loss in energy due to R292K for the weakest inhibitor S00 = 2.07 kcal/mol, which is less than either S01 or S02. This may be related to the suggestion that NA ligands closest in structure to the native substrate would in general be more robust to mutation.[1,80] The pyranose scaffold of S00 is saturated and most like the sialic acid moieties cleaved by NA.

Additional information about the origins of resistance becomes available when considering the differential (relative change) in Coulombic energy. Figure 9b shows the reduced per-residue Coulombic energies ($\Delta\Delta E_{\text{coul}}$) for R292K minus wildtype. Differences evident in the delta footprints reveal some enhanced interactions for S00, particularly at position D151, and R224 which probably contribute to the phenomena that the ligand is second-most robust to the mutation (Figure 9b, red solid line). The most robust inhibitor S03 shows the overall flattest profile (Figure 9b, green dashed line). Most

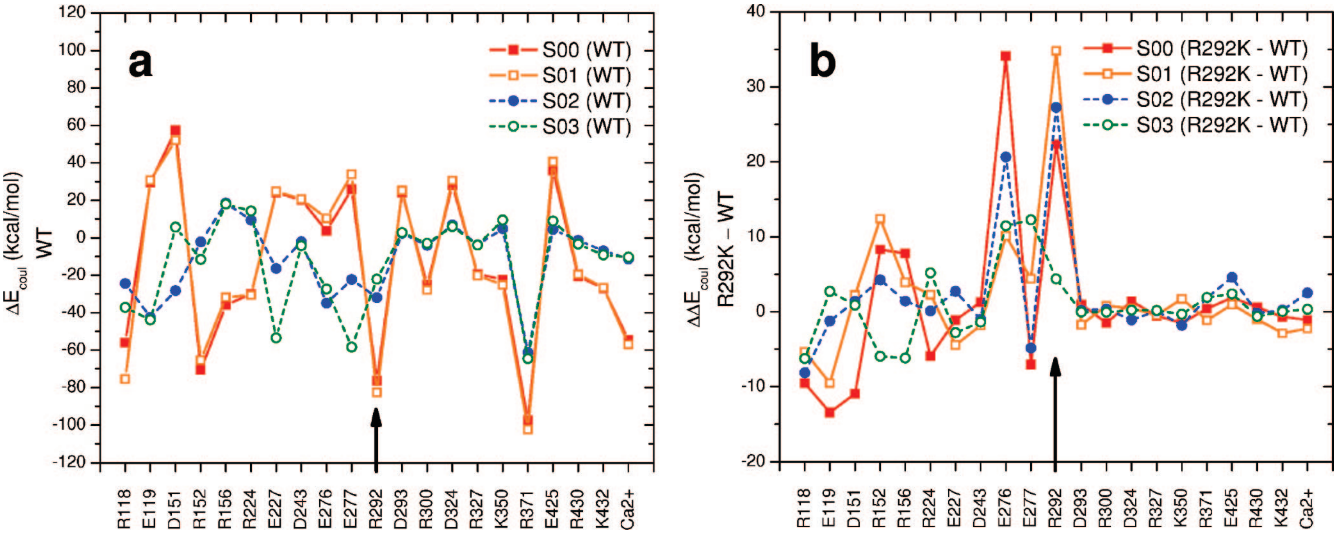


Figure 9. Intermolecular ligand-protein (per-residue) Coulombic energy footprints from a reduced set (favorable or unfavorable with $\Delta E_{\mathrm{coul}} \geq 20$ kcal/mol). **(a)** shows the wildtype NA footprint, while (**b**) shows the difference footprint for the mutation (R292K-WT). Black arrow indicates site of the R292K mutation. Energies in kcal/mol.

significant is the fact that the change at position R292 is relatively constant for S03 compared with the other ligands (Figure 9b, black arrow). Further, the ΔE_{coul} losses here at 292 (energies in kcal/mol) for the series S03 (4.4) < S00 (22.4) < S02 (27.3) < S01 (34.8) parallel the trend for changes in experimental fold resistance (Table 1, $\Delta\Delta G_{\mathrm{R292K\text{-}WT}}$) with S03 (1.66) < S00 (2.07) < S02 (2.70) < S01 (2.76) including the fact that S00 is second-most robust.

Compounds S02 and S03 were originally designed[1,8] using structure-based calculations (GRID program)[33] which predicted that replacement of the -OH group on S01 with positively charged functionality at position C4 on the ligand would lead to enhanced binding due to enhanced electrostatic interactions with glutamic acid at position E119 for S02 and E119 and E227 for S03. Subsequent experimental testing confirmed enhanced binding for S02 and S03 versus S01, and this is often cited as one of the premier examples of rational, computer-aided drug design. The binding footprints computed here (see Figure 9a) clearly illustrate that, as originally predicted,[8] S02 and S03 show enhanced interactions at positions E119 and that S03 makes additional favorable interactions with E227. We also note that additional favorable interactions relative to the other ligands are observed here for S03 with residue E277. As described below, the same general trends observed here for ΔE_{coul} are also seen in H-bonding footprints.

The dynamic nature of the simulations yields protein−ligand interactions with a wide-range of instantaneous per-residue energies, and such variance is expected to be biologically relevant. When viewed as histograms, the electrostatic energies shown in Figure 10 between the ligands and residue 292 only (wildtype Arg or mutant Lys) clearly show that S03 remains relatively invariant to the R292K mutation. Conversely, S00−S02 show dramatic reductions in energy and an overall shift in their respective Coulombic populations (N = 2000). Similarities in magnitude for energies contained in the histograms for charged (S00, S01) versus neutral (S02, S03) ligands are also noted. Overall, the ΔE_{coul} and $\Delta\Delta E_{\mathrm{coul}}$ footprints (Figures 10–12) reveal binding differences which suggest that improved resistance is a function of less reliance on R292 for intrinsic binding affinity and enhanced binding to wildtype via specific interactions primarily at E119, E227, and E277.

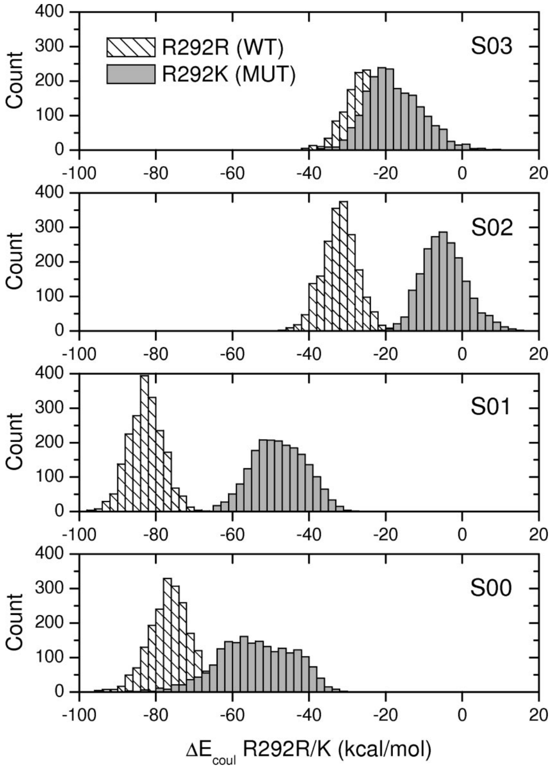


Figure 10. Populations (N = 2000) of per-residue Coulombic energies for ligands with residue 292 for wildtype (hashed histograms) and the R292K mutant (shaded histograms). Energies in kcal/mol.

Hydrogen Bonding. Given the importance that electrostatics play in this system, intermolecular hydrogen bonding was also monitored. The computed values are observed to strongly correlate with experiment. For example, average number of H-bonds from trajectories with wildtype N9 yield S00 (11.46) < S01 (12.10) < S02 (12.95) < S03 (14.23)

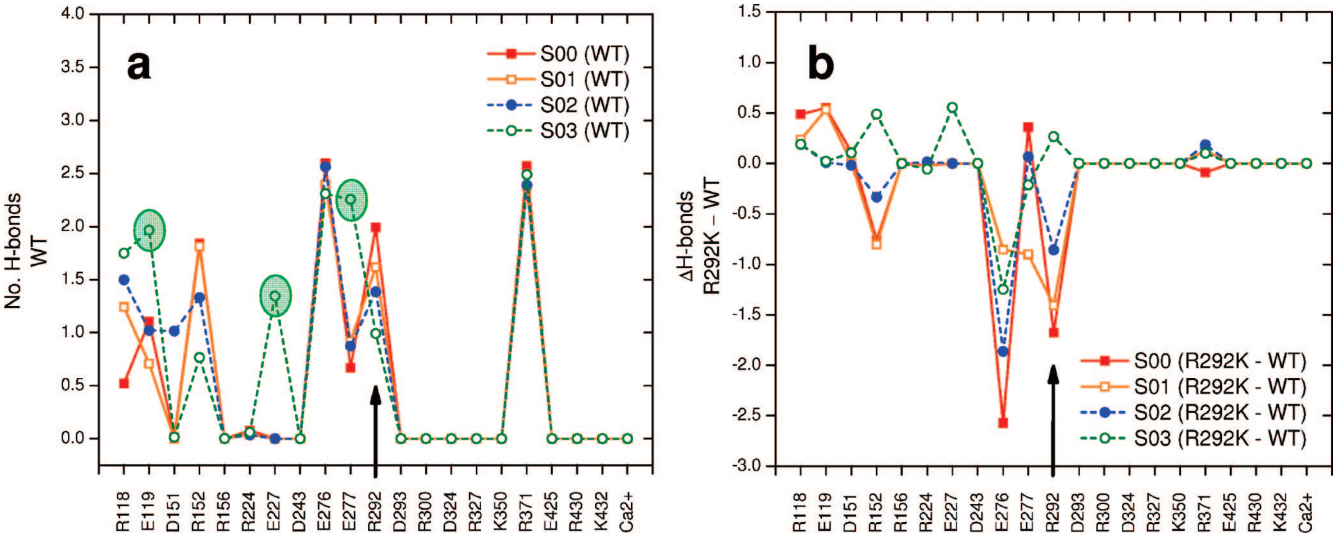


Figure 11. Hydrogen bond footprints (reduced set as in Figure 8) for ligands with NA. **(a)** shows the wildtype NA footprint, while **(b)** shows the difference footprint for the mutation (R292K - WT).

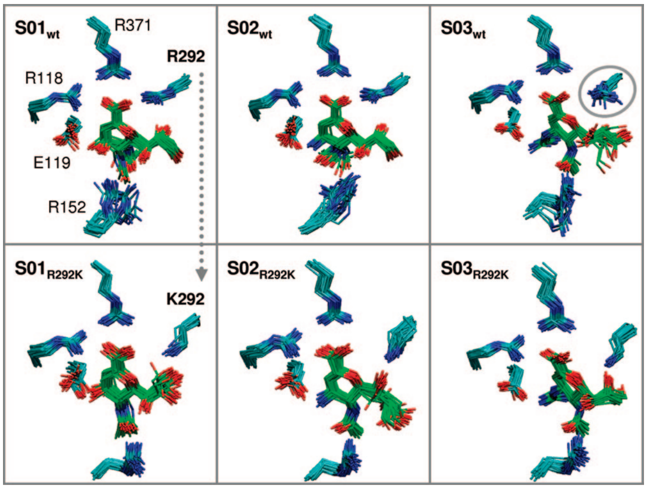


Figure 12. Comparison of snapshots from the MD trajectories of ligand S01–S03 with wildtype NA (**top**) versus the R292K mutation (**bottom**). Forty evenly spaced coordinates sets are presented with only select binding site residues shown for clarity. The arrow (left) indicates the mutation site, and the circled residue (top right panel) highlights the greater flexibility of 292 in the WT complex with S03. Ligand carbon atoms are colored green.

which parallel the binding energies order. And, H-bond values from the R292K simulations yield S00 (7.81) < S01 (8.36) < S02 (9.76) < S03 (15.13) which also follow the experimental ordering (see Table 1). The overall correlation for number of H-bonds with ΔG_{bind} exptl is $r = -0.96$, $r^2 = 0.86$. Interestingly, for R292K, a slight increase is noted for S03 compared with the dramatic losses experienced overall by the other inhibitors and is discussed further below.

Residue-based footprints were also computed to examine if changes in binding energy were structurally correlated to patterns in H-bonds. Figure 11a shows the average number of H-bonds each ligand makes with the reduced set of key pocket residues for wildtype N9. As evident in Figure 11a, all ligands are tightly coordinated to the central Site I Arg at position 371 which makes bidentate interactions with the ligand carboxylate of nearly constant 2.5 H-bonds across the series. These results are consistent with the overall architecture of the NA site (see Figure 1) and mirror the Coulombic energy trends regarding the importance of the Site I central Arg. As was similarly observed for ΔE_{coul} footprints shown in Figure 9a, H-bonding for S03 with R292 is also substantially weaker (H-bonds ca. 1.0) than for the other ligands (H-bonds >1.0) which are all more affected by the mutation. In fact, H-bonds populations at position R292 are inversely correlated to the experimental binding trends with S03 < S02 < S01 < S00 (Figure 11a, black arrow). Again, the implication is that S03 relies less on interaction with R292 for intrinsic affinity with wildtype despite the fact that the ligand is the most potent. Increased potency for S03 is a function of the dramatic increase in H-bonding with a trio of glutamic acids (E119, E227, and E277) as shown in Figure 11a (green ovals). Similar to the ΔE_{coul} profiles (Figure 9), enhanced H-bonding is a result of the much larger guanidinium group interacting with the charged glutamic acids versus smaller C4 functionality (see Table 1) for the other ligands.

Differences in per-residue hydrogen bonding (ΔH-bonds) due to R292K are shown in Figure 11b. Significant numbers of H-bonds are lost for S00–S02 and occur primarily at positions R152, E276, and R292K. At the latter position, losses of 0.9 to 1.7 interactions are observed for inhibitors S00–S02 in contrast with S03 which shows a slight gain (+0.3). The total change in intermolecular ΔH-bonds is S00 (−3.65), S01 (−3.74), S02 (−3.19), and S03 (+0.90). In general, the ΔH-bond profile in Figure 11b for inhibitor S03 is more flat compared with the other ligands as was similarly observed for the $\Delta\Delta E_{\text{coul}}$ profile shown in Figure 9.

A visual examination of MD trajectories reveals that the S03 glycerol group becomes more ordered and locked in place as a result of the mutation compared with ligands S01 and S02 which show the opposite effect (Figure 12). For S03, the R292K side chain also becomes ordered, and this probably coincides with the slight overall increase in H-bonds

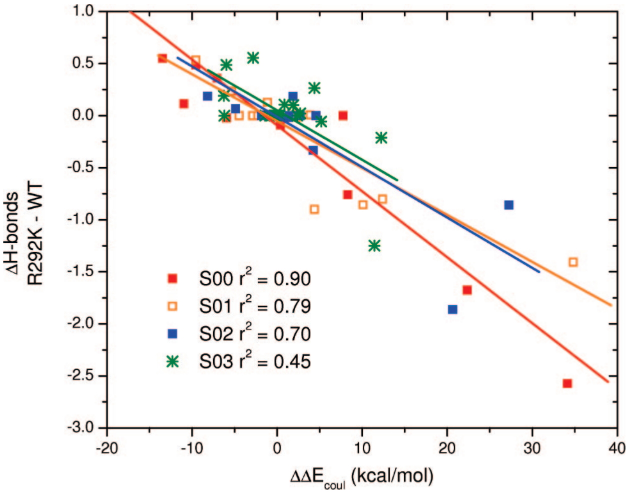


Figure 13. Correlation coefficients (r^2 values) computed between two binding site footprints, ΔH-bonds and $\Delta\Delta E_{coul}$, using the reduced set of binding site residues ($N = 21$).

for S03 as seen in Figure 11b at select positions. Interestingly, Arg 292 in wildtype simulations for S03 shows more disorder vs other ligands (Figure 12, circled residue top right panel). Greater motion here would likely lead to less H-bonds occurring with S03 at this position as is observed in Figure 11a (black arrow). And, as a consequence, S03 is expected to have less reliance on R292 for binding. Distance calculations from wildtype simulations support this hypothesis; average ligand carboxylate carbon to R292@CZ distances are larger for S03 (4.7 Å) versus S00, S01, and S02 (4.3 to 4.4 Å). At the same time, shorter distances are observed between S03 ligand scaffold atoms at position C4 (see Table 1 for numbering) and protein carboxylate atoms at position CD for residues E227 (6.0 vs 7.0 − 8.1 Å) and E277 (3.9 vs 5.4 − 5.8 Å). Here too, shortened distances are a likely consequence of stronger ΔE_{coul} (Figure 9a) and H-bonding (Figure 11a) interactions involving the S03 guanidinium group with E227 and E277.

Footprint Correlation. Finally, the similarity in some of the trends for Coulombic interaction energies and number of H-bonds suggests there may be a quantitative correlation between these descriptors. Figure 13 shows plots of ΔH-bonds vs $\Delta\Delta E_{coul}$ using the per-residue difference footprints (R292K-WT) shown in Figures 11b and 13b. Interestingly, the computed correlation coefficients (r^2 values) between these two terms inversely follow the absolute experimental binding energies with the weakest binder S00 showing the strongest correlation ($r^2 = 0.90$) followed in turn by S01 ($r^2 = 0.79$), S02 ($r^2 = 0.70$), and S03 ($r^2 = 0.45$). Inhibitors observed to lose both Coulombic energy and H-bonding at specific residues (high r^2 values) appear to rely most strongly on those residues for overall intrinsic binding versus those with flatter difference footprints (i.e., S02 and S03).

A comparison of the footprint r^2 values in Figure 13 with changes in experimental fold resistance ($\Delta\Delta G_{R292K\text{-}WT}$, Table 1) reveals a similar trend for S01 ($r^2 = 0.79$, 2.76 kcal/mol) > S02 ($r^2 = 0.70$, 2.70 kcal/mol) > S03 ($r^2 = 0.44$, 1.66 kcal/mol). However, since the weakest binder S00 actually has the second-best fold resistance value (2.07 kcal/mol), inclusion here in this pattern is not expected. Overall, the accord between r^2 values from plots of ΔH-bonds vs $\Delta\Delta E_{coul}$ with the experimental ordering (both ΔG_{bind} and $\Delta\Delta G_{bind}$) reinforces our observation that electrostatics in general is the best descriptor for understanding variation in affinity for sialic acid-based ligands with neuraminidase.

Conclusion

In this study we have employed explicit solvent all-atom MD simulations, free energy calculations, and per-residue footprint analysis to compute relative binding affinities for sialic acid-based ligands with wildtype neuraminidase subtype N9 and with a R292K mutant. The overall goal is to characterize and delineate specific origins of drug resistance. Convergence and stability of the simulations was carefully monitored through examination of instantaneous structural and energetic properties including instantaneous computed free energies of binding (Figure 6), rmsd values from starting structures, changes in desolvation energy $\Delta\Delta G_{hyd}$, and intermolecular energy components ΔE_{coul} and ΔE_{vdw} (Figure 3). Computed standard errors-of-the-mean are low (Table 2), and all measures suggest that the simulations are reasonably converged and well-behaved (Figures 3 and 6).

Notably, the MM-GBSA affinities show strong correlations with experiment (Figure 5, $r^2 = 0.76$) and in every case correctly show that loss of binding occurs as a result of the R292K mutation (Figure 6). In marked contrast to other systems,[40,41] an examination here of binding components reveal that Coulombic (ΔE_{coul}), as opposed to van der Waals (ΔE_{vdw}) energy, is the best overall descriptor for understanding variation in affinity with structure as evident by the significant correlation with experiment ($r^2 = 0.93$, Figure 7, Table 2). Conversely, ΔE_{vdw} shows little correlation ($r^2 = 0.23$). Overall, our analysis suggests that the strong correlation of ΔE_{coul} with ΔG_{bind} exptl is a consequence of the NA binding site being highly charged (Figure 1). Despite large desolvation penalties, the negative sum obtained for the $\Delta G_{electro}$ term ($\Delta E_{coul} + \Delta\Delta G_{polar}$) is in general larger in magnitude than ΔE_{vdw} (Table 2) suggesting that electrostatics is the primary driving force for association. The good correlation obtained between $\Delta G_{electro}$ and ΔG_{bind} exptl ($r^2 = 0.73$) reinforces this view.

Given the strong correlation of Coulombic energy with experiment, origins of resistance were examined through per-residue decomposition of ΔE_{coul} and H-bonding for both wildtype and difference (R292K-WT) footprints. Residue-based decomposition (Figures 11–13) reveals that the most potent ligands have (1) less reliance on R292 for intrinsic binding affinity, (2) enhanced binding via E119, E227, and E277, and (3) flatter overall ΔE_{coul} and ΔH-bond profiles. Wildtype Coulombic and H-bonding footprints confirm the importance of the Site I central Arg at position R371 for all ligands (Figures 11 and 13) and importantly reveal that S03 makes substantially weaker ΔE_{coul} (Figure 9) and H-bond (Figure 11) interactions with residue R292. Weaker interactions with R292 likely contribute to the fact that a mutation at this site is less detrimental for S03. Coulombic energy losses localized at position 292 (Figure 9b, $\Delta\Delta E_{coul}$) nicely parallel the trend for changes in experimental fold resistance energy (Table 2, $\Delta\Delta G_{R292K\text{-}WT}$) with S03 < S00 < S02 <

S01 (Table 1) including the fact S00 is second-most robust to the mutation. Conversely, more favorable interactions for S03 are observed at specific sites both in footprints for ΔE_{coul} (Figure 9a, residues E227, E277) and in footprints for numbers of H-bonds (Figure 11a, residues E119, E227, E277). Stronger per-residue interactions at these positions are a function of S03's larger guanidinium group.

Total H-bond populations for wildtype and R292K also parallel the experimental ordering ($r = -0.96$, $r^2 = 0.86$). Despite the fact S03 makes the overall largest number of H-bonds with NA, counts localized at position 292 are actually inversely correlated with experiment (Figure 11a) again suggesting S03 relies less on R292 for wildtype affinity. Upon mutation, significant numbers of total H-bonds are lost for ligands S00 (−3.26), S01 (−3.74), and S02 (−3.19), in particular at positions R152, E276, and R292K (Figure 11b). However, for S03 a slight increase in H-bonding (+0.90) is observed which can in part be traced to the fact that both residue 292 and the ligand glycerol side chain both become slightly more ordered (Figure 12). Additionally, wildtype simulations show that Arg at 292 is more disordered in simulations with S03 which helps to explain why S03 shows less H-bonding at this position (Figure 11a). Supporting this hypothesis, ligand carboxylate C1 distances with R292@CZ are longer for S03 than other inhibitors (4.7 vs 4.3 − 4.4 Å), while ligand scaffold C4 distances to E227@CD (6.0 vs 7.0 − 8.1 Å) and E277@CD (3.9 vs 5.4 − 5.8 Å) are shorter. The net result is weaker interactions for S03 with R292 and a more robust resistance profile.

The highly variable nature of the influenza virus, combined with the possibility for interspecies infection and transmission, represents a major challenge for development of both timely vaccines and development of robust antivirals active against various strains and subtypes. For this reason, studies geared toward characterization and increased understanding of how ligands bind with their antiviral targets are paramount. In this study, we have participated toward this goal by demonstrating that all-atom computer simulations with energetic and structural analysis, for a series of ligands with neuraminidase from influenza subtype N9, yield computed free energies of binding that agree well with experiment. In particular, our simulations correctly predict the effects of a known mutation at position R292K and provide clues as to origins of resistance to the mutant. Use of residue-based decomposition highlights the power of computational methods for probing specific binding interfaces and for characterization of which specific interactions govern molecular recognition. Overall, the results significantly enhance experimental observations. The likelihood of future influenza pandemics (including the possibility of highly pathogenic H5N1 strains) highlights the need for additional computational modeling studies that continue to address binding and origins of drug resistance.

Acknowledgment. Gratitude is expressed to Noel Carrascal and Rashi Goyal for computational assistance, to Janet Hearing for helpful discussions, and to the New York State Office of Science Technology and Academic Research (NYSTAR), the Office of the Vice President of Research at Stony Brook University, the Computational Science Center at Brookhaven National Laboratory, and the Stony Brook School of Medicine Targeted Research Opportunity Program for support of this research. R.C. was supported by Undergraduate Fellowships from HHMI (Award Number 52003746) and the Stony Brook URECA Program.

References

(1) von Itzstein, M. The war against influenza: discovery and development of sialidase inhibitors. *Nat. Rev. Drug Discovery* **2007**, *6*, 967–74.

(2) A revision of the system of nomenclature for influenza viruses: a WHO memorandum. *Bull. World Health Organ.* **1980**, *58*, 585–91.

(3) WHO Epidemic and Pandemic Alert and Response (EPR): Avian Influenza. http://www.who.int/csr/disease/influenza (accessed June 1, 2008).

(4) Nichol, K. L.; Treanor, J. J. Vaccines for seasonal and pandemic influenza. *J. Infect. Dis.* **2006**, *194 Suppl 2*, S111−8.

(5) Johnson, N. P.; Mueller, J. Updating the accounts: global mortality of the 1918−1920 "Spanish" influenza pandemic. *Bull. Hist. Med.* **2002**, *76*, 105–15.

(6) Rajagopal, S.; Treanor, J. Pandemic (avian) influenza. *Semin. Respir. Crit. Care Med.* **2007**, *28*, 159–70.

(7) Subbarao, K.; Klimov, A.; Katz, J.; Regnery, H.; Lim, W.; Hall, H.; Perdue, M.; Swayne, D.; Bender, C.; Huang, J.; Hemphill, M.; Rowe, T.; Shaw, M.; Xu, X.; Fukuda, K.; Cox, N. Characterization of an avian influenza A (H5N1) virus isolated from a child with a fatal respiratory illness. *Science* **1998**, *279*, 393–6.

(8) von Itzstein, M.; Wu, W. Y.; Kok, G. B.; Pegg, M. S.; Dyason, J. C.; Jin, B.; Van Phan, T.; Smythe, M. L.; White, H. F.; Oliver, S. W.; et al. Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature* **1993**, *363*, 418–23.

(9) Taylor, N. R.; von Itzstein, M. Molecular modeling studies on ligand binding to sialidase from influenza virus and the mechanism of catalysis. *J. Med. Chem.* **1994**, *37*, 616–24.

(10) Jedrzejas, M. J.; Singh, S.; Brouillette, W. J.; Air, G. M.; Luo, M. A strategy for theoretical binding constant, Ki, calculations for neuraminidase aromatic inhibitors designed on the basis of the active site structure of influenza virus neuraminidase. *Proteins: Struct., Funct., Genet.* **1995**, *23*, 264–77.

(11) Taylor, N. R.; von Itzstein, M. A structural and energetics analysis of the binding of a series of N-acetylneuraminic-acid-based inhibitors to influenza virus sialidase. *J. Comput.-Aided Mol. Des.* **1996**, *10*, 233–246.

(12) Wall, I. D.; Leach, A. R.; Salt, D. W.; Ford, M. G.; Essex, J. W. Binding constants of neuraminidase inhibitors: An investigation of the linear interaction energy method. *J. Med. Chem.* **1999**, *42*, 5142–52.

(13) Muegge, I. The effect of small changes in protein structure on predicted binding modes of known inhibitors of influenza virus neuraminidase: PMF-scoring in DOCK4. *Med. Chem. Res.* **1999**, *9*, 490–500.

(14) Wang, T.; Wade, R. C. Comparative binding energy (COMBINE) analysis of influenza neuraminidase-inhibitor complexes. *J. Med. Chem.* **2001**, *44*, 961–71.

(15) Smith, B. J.; McKimm-Breshkin, J. L.; McDonald, M.; Fernley, R. T.; Varghese, J. N.; Colman, P. M. Structural studies of the resistance of influenza virus neuramindase to inhibitors. *J. Med. Chem.* **2002**, *45*, 2207–12.

(16) Masukawa, K. M.; Kollman, P. A.; Kuntz, I. D. Investigation of Neuraminidase-Substrate Recognition Using Molecular Dynamics and Free Energy Calculations. *J. Med. Chem.* **2003**, *46*, 5628–5637.

(17) Bonnet, P.; Bryce, R. A. Molecular dynamics and free energy analysis of neuraminidase-ligand interactions. *Protein Sci.* **2004**, *13*, 946–957.

(18) Bonnet, P.; Bryce, R. A. Scoring binding affinity of multiple ligands using implicit solvent and a single molecular dynamics trajectory: Application to Influenza neuraminidase. *J. Mol. Graphics Modell.* **2005**, *24*, 147–156.

(19) Verma, R. P.; Hansch, C. A QSAR study on influenza neuraminidase inhibitors. *Bioorg. Med. Chem.* **2006**, *14*, 982–96.

(20) Amaro, R. E.; Minh, D. D.; Cheng, L. S.; Lindstrom, W. M., Jr.; Olson, A. J.; Lin, J. H.; Li, W. W.; McCammon, J. A. Remarkable loop flexibility in avian influenza N1 and its implications for antiviral drug design. *J. Am. Chem. Soc.* **2007**, *129*, 7764–5.

(21) NIH Fact Sheet. NIGMS-Supported Structure-Based Drug Design Saves Lives. http://www.nigms.nih.gov/Publications/structure_drugs.htm (accessed June 1, 2008).

(22) FDA Approved Drugs. http://www.fda.gov/cder/drug/antivirals/influenza/default.htm#drugs (accessed June 1, 2008).

(23) BioCryst Pharmaceuticals. http://www.biocryst.com/peramivir.htm (accessed June 1, 2008).

(24) Varghese, J. N.; Laver, W. G.; Colman, P. M. Structure of the influenza virus glycoprotein antigen neuraminidase at 2.9 A resolution. *Nature* **1983**, *303*, 35–40.

(25) Stoll, V.; Stewart, K. D.; Maring, C. J.; Muchmore, S.; Giranda, V.; Gu, Y. G. Y.; Wang, G.; Chen, Y. W.; Sun, M. H.; Zhao, C.; Kennedy, A. L.; Madigan, D. L.; Xu, Y. B.; Saldivar, A.; Kati, W.; Laver, G.; Sowin, T.; Sham, H. L.; Greer, J.; Kempf, D. Influenza neuraminidase inhibitors: Structure-based design of a novel inhibitor series. *Biochemistry* **2003**, *42*, 718–727.

(26) McKimm-Breschkin, J. L.; Sahasrabudhe, A.; Blick, T. J.; McDonald, M.; Colman, P. M.; Hart, G. J.; Bethell, R. C.; Varghese, J. N. Mutations in a conserved residue in the influenza virus neuraminidase active site decreases sensitivity to Neu5Ac2en- derived inhibitors. *J. Virol.* **1998**, *72*, 2456–2462.

(27) Blick, T. J.; Tiong, T.; Sahasrabudhe, A.; Varghese, J. N.; Colman, P. M.; Hart, G. J.; Bethell, R. C.; McKimm-Breschkin, J. L. Generation and characterization of an influenza virus neuraminidase variant with decreased sensitivity to the neuraminidase-specific inhibitor 4-guanidino-Neu5Ac2en. *Virology* **1995**, *214*, 475–84.

(28) Le, Q. M.; Kiso, M.; Someya, K.; Sakai, Y. T.; Nguyen, T. H.; Nguyen, K. H.; Pham, N. D.; Ngyen, H. H.; Yamada, S.; Muramoto, Y.; Horimoto, T.; Takada, A.; Goto, H.; Suzuki, T.; Suzuki, Y.; Kawaoka, Y. Avian flu: isolation of drug-resistant H5N1 virus. *Nature* **2005**, *437*, 1108.

(29) Abed, Y.; Nehme, B.; Baz, M.; Boivin, G. Activity of the neuraminidase inhibitor A-315675 against oseltamivir-resistant influenza neuraminidases of N1 and N2 subtypes. *Antiviral Res.* **2008**, *77*, 163–6.

(30) McKimm-Breschkin, J. L. Resistance of influenza viruses to neuraminidase inhibitors—a review. *Antiviral Res.* **2000**, *47*, 1–17.

(31) Zambon, M.; Hayden, F. G. Position statement: global neuraminidase inhibitor susceptibility network. *Antiviral Res.* **2001**, *49*, 147–56.

(32) Reece, P. A. Neuraminidase inhibitor resistance in influenza viruses. *J. Med. Virol.* **2007**, *79*, 1577–86.

(33) Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **1985**, *28*, 849–57.

(34) Armstrong, K. A.; Tidor, B.; Cheng, A. C. Optimal charges in lead progression: a structure-based neuraminidase case study. *J. Med. Chem.* **2006**, *49*, 2470–7.

(35) Russell, R. J.; Haire, L. F.; Stevens, D. J.; Collins, P. J.; Lin, Y. P.; Blackburn, G. M.; Hay, A. J.; Gamblin, S. J.; Skehel, J. J. The structure of H5N1 avian influenza neuraminidase suggests new opportunities for drug design. *Nature* **2006**, *443*, 45–9.

(36) De Clercq, E. Antiviral agents active against influenza A viruses. *Nat. Rev. Drug Discovery* **2006**, *5*, 1015–25.

(37) Moscona, A. Neuraminidase inhibitors for influenza. *N. Engl. J. Med.* **2005**, *353*, 1363–73.

(38) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.* **2000**, *33*, 889–97.

(39) Massova, I.; Kollman, P. A. Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding. *Perspect. Drug Discovery Des.* **2000**, *18*, 113–135.

(40) Strockbine, B.; Rizzo, R. C. Binding of Anti-fusion Peptides with HIVgp41 from Molecular Dynamics Simulations: Quantitative Correlation with Experiment. *Proteins: Struct., Funct., Bioinf.* **2007**, *63*, 630–642.

(41) Rizzo, R. C.; Toba, S.; Kuntz, I. D. A Molecular Basis for the Selectivity of Thiadiazole Urea Inhibitors with Stromelysin-1 and Gelatinase-A from Generalized Born Molecular Dynamics Simulations. *J. Med. Chem.* **2004**, *47*, 3065–3074.

(42) Onufriev, A.; Bashford, D.; Case, D. A. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins: Struct., Funct., Bioinf.* **2004**, *55*, 383–94.

(43) Feig, M.; Onufriev, A.; Lee, M. S.; Im, W.; Case, D. A.; Brooks, C. L., III. Performance comparison of generalized born and Poisson methods in the calculation of electrostatic solvation energies for protein structures. *J. Comput. Chem.* **2004**, *25*, 265–84.

(44) Rizzo, R. C.; Aynechi, T.; Case, D. A.; Kuntz, I. D. Estimation of Absolute Free Energies of Hydration Using Continuum Methods: Accuracy of Partial Charge Models and Optimization of Nonpolar Contributions. *J. Chem. Theory Comput.* **2006**, *2*, 128–139.

(45) Jorgensen, W. L. Free Energy Calculations: A Breakthrough for Modeling Organic Chemistry in Solution. *Acc. Chem. Res.* **1989**, *22*, 184–189.

(46) Kollman, P. Free Energy Calculations: Applications to Chemical and Biochemical Phenomena. *Chem. Rev.* **1993**, *93*, 2395–2417.

(47) Kormos, B. L.; Benitex, Y.; Baranger, A. M.; Beveridge, D. L. Affinity and specificity of protein U1A-RNA complex formation based on an additive component free energy model. *J. Mol. Biol.* **2007**, *371*, 1405–19.

(48) Lyne, P. D.; Lamb, M. L.; Saeh, J. C. Accurate prediction of the relative potencies of members of a series of kinase inhibitors using molecular docking and MM-GBSA scoring. *J. Med. Chem.* **2006**, *49*, 4805–8.

(49) Gohlke, H.; Case, D. A. Converging free energy estimates: MM-PB(GB)SA studies on the protein-protein complex Ras-Raf. *J. Comput. Chem.* **2004**, *25*, 238–250.

(50) 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.

(51) *MOE*; Chemical Computing Group: Montreal, Canada, 2007.

(52) Smith, B. J.; Colman, P. M.; Von Itzstein, M.; Danylec, B.; Varghese, J. N. Analysis of inhibitor binding in influenza virus neuraminidase. *Protein Sci.* **2001**, *10*, 689–696.

(53) *AMBER Version 8*; University of California at San Francisco: San Francisco, CA, 2004.

(54) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–74.

(55) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 712–25.

(56) Fornabaio, M.; Cozzini, P.; Mozzarelli, A.; Abraham, D. J.; Kellogg, G. E. Simple, intuitive calculations of free energy of binding for protein-ligand complexes. 2. Computational titration and pH effects in molecular models of neuraminidase-inhibitor complexes. *J. Med. Chem.* **2003**, *46*, 4487–500.

(57) 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.

(58) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G. ; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98, revision A.9*; Gaussian Inc.: Pittsburgh, PA, 1998.

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

(60) Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. Molecular-Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

(61) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.

(62) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

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

(64) Tsui, V.; Case, D. A. Theory and applications of the generalized Born solvation model in macromolecular simulations. *Biopolymers* **2000**, *56*, 275–91.

(65) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Pairwise Solute Descreening of Solute Charges from a Dielectric Medium. *Chem. Phys. Lett.* **1995**, *246*, 122–129.

(66) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. *J. Phys. Chem.* **1996**, *100*, 19824–19839.

(67) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127–6129.

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

(69) Sitkoff, D.; Sharp, K. A.; Honig, B. Accurate Calculation of Hydration Free-Energies Using Macroscopic Solvent Models. *J. Phys. Chem.* **1994**, *98*, 1978–1988.

(70) Lee, M. R.; Duan, Y.; Kollman, P. A. Use of MM-PB/SA in estimating the free energies of proteins: application to native, intermediates, and unfolded villin headpiece. *Proteins: Struct., Funct., Genet.* **2000**, *39*, 309–16.

(71) Vorobjev, Y. N.; Hermans, J. ES/IS: estimation of conformational free energy by combining dynamics simulations with explicit solvent with an implicit solvent continuum model. *Biophys. Chem.* **1999**, *78*, 195–205.

(72) Vorobjev, Y. N.; Almagro, J. C.; Hermans, J. Discrimination between native and intentionally misfolded conformations of proteins: ES/IS, a new method for calculating conformational free energy that uses both dynamics simulations with an explicit solvent and an implicit solvent continuum model. *Proteins: Struct., Funct., Genet.* **1998**, *32*, 399–413.

(73) Morozov, A. V.; Kortemme, T.; Baker, D. Evaluation of Models of Electrostatic Interactions in Proteins. *J. Phys. Chem. B* **2003**, *107*, 2075–2090.

(74) Rizzo, R. C.; Tirado-Rives, J.; Jorgensen, W. L. Estimation of binding affinities for HEPT and nevirapine analogues with HTV-1 reverse transcriptase via Monte Carlo simulations. *J. Med. Chem.* **2001**, *44*, 145–154.

(75) Rizzo, R. C.; Udier-Blagovic, M.; Wang, D. P.; Watkins, E. K.; Smith, M. B. K.; Smith, R. H.; Tirado-Rives, J.; Jorgensen, W. L. Prediction of activity for nonnucleoside inhibitors with HIV-1 reverse transcriptase based on Monte Carlo simulations. *J. Med. Chem.* **2002**, *45*, 2970–2987.

(76) Rafi, S. B.; Cui, G.; Song, K.; Cheng, X.; Tonge, P. J.; Simmerling, C. Insight through molecular mechanics Poisson-Boltzmann surface area calculations into the binding affinity of triclosan and three analogues for FabI, the E. coli enoyl reductase. *J. Med. Chem.* **2006**, *49*, 4574–80.

(77) *DOCK*; University of California at San Francisco: San Francisco, CA, 2007.

(78) Ewing, T. J.; Makino, S.; Skillman, A. G.; Kuntz, I. D. DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 411–28.

(79) Moustakas, D. T.; Therese Lang, P. T.; Pegg, S.; Pettersen, E.; Kuntz, I. D.; Broojimans, N.; Rizzo, R. C. Development and Validation of a Modular, Extensible Docking program: DOCK 5. *J. Comput.-Aided Mol. Des.* **2006**, *20*, 601−619.

(80) Varghese, J. N.; Smith, P. W.; Sollis, S. L.; Blick, T. J.; Sahasrabudhe, A.; McKimm-Breschkin, J. L.; Colman, P. M. Drug design against a shifting target: a structural basis for resistance to inhibitors in a variant of influenza virus neuraminidase. *Struct. Fold. Des.* **1998**, *6*, 735–746.

CT800068V

JCTC Journal of Chemical Theory and Computation

Intrinsic Free Energy of the Conformational Transition of the KcsA Signature Peptide from Conducting to Nonconducting State

Ilja V. Khavrutskii,*,[†] Mikolai Fajer,[†] and J. Andrew McCammon[†,‡]

Howard Hughes Medical Institute, Center for Theoretical Biological Physics,
Departments of Chemistry and Biochemistry and Pharmacology, University of
California San Diego, La Jolla, California 92093-0365

Received March 11, 2008

Abstract: We explore a conformational transition of the TATTVGYG signature peptide of the KcsA ion selectivity filter and its GYG to AYA mutant from the conducting α-strand state into the nonconducting pII-like state using a novel technique for multidimensional optimization of transition path ensembles and free energy calculations. We find that the wild type peptide, unlike the mutant, intrinsically favors the conducting state due to G77 backbone propensities and additional hydrophobic interaction between the V76 and Y78 side chains in water. The molecular mechanical free energy profiles in explicit water are in very good agreement with the corresponding adiabatic energies from the Generalized Born Molecular Volume (GBMV) implicit solvent model. However comparisons of the energies to higher level B3LYP/6−31G(d) Density Functional Theory calculations with Polarizable Continuum Model (PCM) suggest that the nonconducting state might be more favorable than predicted by molecular mechanics simulations. By extrapolating the single peptide results to the tetrameric channel, we propose a novel hypothesis for the ion selectivity mechanism.

Introduction

Organisms transmit electric impulses by means of cellular membrane polarization that critically depends on the work of ion channels. These channels permit passage of specific ion types across the membrane. Ion channels selective for potassium such as KcsA[1,2] are particularly interesting as they solve a nontrivial problem of selecting larger K^+ over smaller Na^+ ions. Despite the wealth of information derived from both experimental[1–4] and computational[5–18] studies of potassium channels, the mechanism of selectivity in these biological machines remains too difficult to tackle as it requires probing the multi-ion permeation transition states.[3–5,7,8,11–18] From the computational perspective, this task demands computing multidimensional potentials of mean force (PMFs) for which efficient tools have been lacking.[19–36]

Recently, we have developed and generalized the gradient-augmented Harmonic Fourier Beads (ggaHFB) method[21–23] that allows studying rare events in complex molecular systems by extending Fukui's intrinsic reaction coordinate (IRC) approach[37,38] with the help of the multidimensional free-energy gradient.[22,23,39,40]

In the present paper we apply the ggaHFB methodology to study an important functional transition of the signature peptide TATTVGYG of the KcsA selectivity filter that pinches the filter shut by flipping its V76 carbonyl group away from the channel axis coupled with the V76 side chain rotation in response to lowering the K^+ concentration.[1,3,4] The V76 carbonyl group flip in the KcsA channel is associated with the α_L to pII backbone conformational transition at the G77 residue of the signature peptide and is believed to switch the selectivity filter from a conducting (α_L) to a nonconducting (pII) state. This transition has been alluded to by X-ray crystallography that detected a partial

[*] Corresponding author e-mail: ikhavru@mccammon.ucsd.edu.

[†] Department of Chemistry and Biochemistry, University of California San Diego.

[‡] Department of Pharmacology, University of California San Diego.

10.1021/ct800086s CCC: $40.75 © 2008 American Chemical Society
Published on Web 08/13/2008

flip of the V76 backbone carbonyls in the wild type KcsA upon lowering K^+ concentration[1,2] and recently a more pronounced flip in the E71A mutant.[41] Similar transitions have been observed in numerous molecular dynamics (MD) simulations of KcsA[7,8] and other related channels.[42–44] Interestingly, X-ray crystallographic studies indicate that the α_L to pII backbone transition is accompanied by rotation of the V76 side chain. However, to the best of our knowledge, previous MD simulations of the KcsA and related potassium channels did not report such a rotation. Furthermore, while the carbonyl flip into pII state observed by X-ray crystallography preserved the 4-fold symmetry of the channel, the MD simulations reported only a single strand out of four identical strands to undergo the α_L to pII transition, thus breaking the symmetry of the channel.

It is possible that averaging over the four strands of the filter might artificially diminish the extent of the transition seen by the X-ray crystallography, thus masking the symmetry breaking. However, unambiguous demonstration of the symmetry breaking requires assessing the free energy of the conformational transition in the full tetrameric channel. Although possible to accomplish with the help of the ggaHFB method, this task is computationally intensive as it requires free energy optimization of a transition path ensemble for a relatively large system. On the other hand, exploring the same transition using a single peptide might provide useful insights into the function of the tetrameric channel with reduced computational burden. In particular, the intrinsic free energy profile should provide relative free energies of the α_L and pII states along with the corresponding free energy barrier outside the channel environment and thus suggest whether multiple transitions inside the channel are likely.

We define the intrinsic free energy profile of the peptide as that of a single peptide in water. Our choice of water medium has been motivated by the following observations. The distributions of the Ramachandran dihedral angles of various residues in the existing protein structures resemble those from the corresponding adiabatic maps in water but differ markedly from those in gas phase.[45–49] Even though KcsA is a trans-membrane protein, when fully assembled and in conducting state, water molecules can access the back of the selectivity filter, where they participate in hydrogen bonding with E71 and D80 residues (not present in our model).[1,2,15,50] Additional water molecules reach behind the selectivity filter to interact with other residues of the signature peptide in the nonconducting state.[2,3] Furthermore, the filter is known to conduct water with and without the ions and hence has a water accessible interior.[1–3,51] Therefore, we feel that the study of the behavior of a single selectivity peptide in water will provide useful insights for understanding the behavior of the same peptide in the tetrameric channel.

This paper is organized as follows. First, we review the ggaHFB methodology for finding minimum adiabatic potential energy paths and minimum free energy transition path ensembles and computing corresponding energy profiles. Combining the ggaHFB transition path ensemble optimization and free-energy evaluation capabilities with the available X-ray structural information, we then explore the intrinsic free energy profile of the signature peptide underlying the flip of the V76 carbonyl from conducting into the nonconducting state.[52–56] Furthermore, we evaluate the effect of the V76 side chain rotation on the backbone transition. To derive additional support for the functional importance of the specified transition to the KcsA channel, we compare the free energy profile of the wild type peptide to that of the GYG to AYA mutant. Note that a closely related G77A mutant either abolishes the selectivity[57] or abrogates the activity of the channel.[58] To diffuse any doubts regarding the choice of the water environment for our study, we examine the changes to the functional transition upon removing the peptide from water and placing it into gas phase. Here we fully utilize the ggaHFB capabilities in finding minimum adiabatic potential energy pathways and computing the corresponding energy profiles via the generalized line integral formalism. Finally, we provide some benchmarks to lend credence to the computed energy profiles in water. In particular, we gauge the molecular mechanical (MM) CHARMM22 force field[59,60] against a popular Quantum Mechanical (QM) Density Functional Theory model, namely B3LYP[61–63] with a 6−31G(d) basis set. To account for the solvent contribution, we employ the Generalized Born Molecular Volume (GBMV)[47,64] and Polarizable Continuum Model (PCM)[65–68] with the MM and QM energy functions, respectively.

Methodology

Given the novelty of the employed transition path and path ensemble optimization technique−the generalized gradient augmented Harmonic Fourier Beads method−that makes this study possible, we briefly describe the main points of the method in the following paragraphs.

1. Reactive Coordinate Space (RCS) and Biasing Potential. The generalized gradient-augmented Harmonic Fourier Beads (ggaHFB) method considers an arbitrary system of N atoms described by $3N$ generalized coordinates $\overline{Q} = (q_1, \cdots, q_{3N})$, and, equivalently, by $3N$ Cartesian coordinates $\overline{X} = (x_1, \cdots, x_{3N})$. The method derives the gradient of either adiabatic potential energy or the free energy of the system with respect to a selected subset of $S \le 3N$ coordinates $\overline{q} = (q_1, \cdots, q_S)$ that comprise the reactive coordinate space (RCS) by employing either biased optimization or biased molecular dynamics (MD) or Monte Carlo (MC) simulations, correspondingly. The remaining $3N$-S degrees of freedom $\overline{r} = (q_{S+1}, \cdots, q_{3N})$ comprise the spectator coordinate space (SCS) and do not contribute explicitly to the energy gradient.

The biasing potential is a linear combination of relatively stiff harmonic restraints and applies only to the RCS degrees of freedom centered at a reference configuration $\overline{q}^{b,ref} = (q_1^{b,ref}, \cdots, q_S^{b,ref})$:[22,23]

$$V^b(q_1, \cdots, q_S; q_1^{b,ref}, \cdots, q_S^{b,ref}) = \sum_{i=1}^{S} k_i^b (q_i - q_i^{b,ref})^2 \tag{1}$$

Here superscript b indicates the bias, and k_i^b is the i^{th} coordinate bias force constant. This biasing potential allows deriving the desired energy gradients using a very simple idea described in the following section.

2. Adiabatic Potential Energy Gradient from Biased Optimization. The key idea for computing the energy gradients is most clearly demonstrated on the example of the adiabatic potential energy. Let us add the biasing potential (1) to the total energy of the system $U(\overline{Q}) = U(\overline{q},\overline{r})$ and then perform potential energy optimization on the modified potential energy surface. Such optimization should reach an equilibrium point at which the forces from the biasing potential that apply only to the S degrees of freedom balance those from the potential energy. Because the forces on the remaining $3N$-S degrees of freedom become identically zero due to optimization, the equilibrium point provides the gradient not of the full potential energy but instead of the adiabatic potential energy. Therefore, the biased optimization yields the gradient of the adiabatic potential energy in RCS. The following equations summarize the above.

$$\left.\frac{\partial U(\overline{q},\overline{r})}{\partial q_i}\right|_{\overline{q}=[\overline{q}]^b,\overline{r}=[\overline{r}]} = -\left.\frac{\partial V(\overline{q})}{\partial q_i}\right|_{\overline{q}=[\overline{q}]^b} = -2k_i^b([q_i]^b - q_i^{b,ref}),\ \text{for } i = 1,\cdots,S$$

$$\left.\frac{\partial U(\overline{q},\overline{r})}{\partial q_i}\right|_{\overline{q}=[\overline{q}]^b,\overline{r}=[\overline{r}]} = 0,\ \text{for } i = S+1,\cdots,3N \qquad (2)$$

The square brackets indicate the local minimum on the modified potential energy surface. This procedure effectively reduces the full potential energy surface of $3N$ degrees of freedom to the adiabatic potential energy surface of $S \le 3N$ degrees of freedom.

It is worth noting that in order to compute the adiabatic potential energy gradient on steep slopes in the vicinity of transition states one has to use somewhat stiff springs. Otherwise the minimum on the modified energy surface will slide downhill close to the corresponding minimum on the full energy surface providing little or no information about the transition state region. This remark also applies to the free energy gradient discussed in the next paragraph.

3. Free Energy Gradient From Biased Simulations. The idea used to derive the gradient of the adiabatic potential energy can be applied to derive the gradient of the free energy from biased simulations. For the proof of this statement we refer the reader to the previous work[22,23,39,40] and only summarize the results here. It has been demonstrated that for somewhat stiff Cartesian restraint (1) with reference configuration $\overline{x}^{b,ref}$ in RCS, one can compute the corresponding Cartesian free energy gradient via eq 3a.[22,23,39,40]

$$\left.\frac{\partial W^u(\overline{x})}{\partial x_i}\right|_{\overline{x}=\langle\overline{x}\rangle^b} \approx -2k_i^b(\langle\overline{x}_i\rangle^b - x_i^{b,ref}) \qquad (3a)$$

Similarly, for the restraint (1) in generalized coordinates centered at $\overline{q}^{b,ref}$ the corresponding free energy gradient is given by eq 3b.[23]

$$\left.\frac{\partial W^u(\overline{q})}{\partial q_i}\right|_{\overline{q}=\langle\overline{q}\rangle^b} \approx -2k_i^b(\langle\overline{q}_i\rangle^b - q_i^{b,ref}) + k_BT\left.\frac{\partial \ln|J(\overline{q})|}{\partial q_i}\right|_{\overline{q}=\langle\overline{q}\rangle^b} \qquad (3b)$$

Here W^u is the unbiased free energy, k_B is the Boltzmann constant, T is the simulation temperature, and $|J(\overline{q})|$ is the ensemble-reduced Jacobian for the transformation from Cartesian to the generalized coordinates. Note that eq 3a is practically identical to eq 2 for the adiabatic potential energy gradient, where the biased ensemble average $\langle\overline{q}\rangle^b = (\langle q_1\rangle^b, \cdots,\langle q_S\rangle^b)$ replaces the local minimum $[\overline{q}]^b$ configuration. The additional logarithmic Jacobian term on the right-hand side of the generalized gradient expression 3b is the consequence of using Cartesian MD or MC propagators with the nonlinear restraints.[23,69] Unlike the case for the adiabatic potential energy gradient, the free energy gradient expression is approximate.

The quality of the free energy gradient depends on the stiffness of the harmonic restraint[22,23] and on the quality of the corresponding configuration averages. To achieve the highest quality, one can either run a single very long simulation or run several short simulations and then combine the results into the cumulative average. We prefer the latter approach for accurate free energy calculations as it allows monitoring convergence of the gradient. Specifically, running P batches of short MD or MC simulations of equal length subject to the restraint (1) provides P sets of averaged coordinates or "evolved beads" $\langle\overline{q}\rangle^{b,j} = (\langle q_1\rangle^{b,j}, \cdots,\langle q_S\rangle^{b,j})$ for a given reference bead, where j is the batch number. These averages could then be easily combined to yield the higher quality cumulative average:

$$\langle\overline{q}\rangle^b = \frac{1}{P}\sum_{j=1}^{P}\langle\overline{q}\rangle^{b,j} \qquad (4)$$

Importantly, the averaged configuration provides the complete free energy gradient in RCS and not just one of its components:

$$\nabla W^u(\overline{q})|_{\langle\overline{q}\rangle^b} = \left(\left.\frac{\partial W^u(\overline{q})}{\partial q_1}\right|_{\langle\overline{q}\rangle^b},\cdots,\left.\frac{\partial W^u(\overline{q})}{\partial q_S}\right|_{\langle\overline{q}\rangle^b}\right) \qquad (5)$$

This property of the ggaHFB method is a great advantage over the histogram-based free energy estimates that require much larger arrays of simulations to populate multidimensional histograms.[22,23,39,40,70,71] Therefore, the ggaHFB method offers a practical alternative to the conventional umbrella sampling simulations with weighted histogram analysis method (WHAM).[70,71]

The ability to compute the free energy gradient efficiently makes it possible to perform gradient-driven optimization on free energy surfaces and ultimately to find minimum free energy transition path ensembles.

4. Minimum Adiabatic Energy Transition Path. The ggaHFB method as a path finding tool belongs to the class of double-ended reaction path methods that require a reactant and a product state to describe a transition of interest.[19–26,72–79] Importantly, the ggaHFB method finds reaction or transition paths that are invariant with respect to coordinate transformations. The concept of invariant reaction paths, called "intrinsic reaction coordinate" (IRC), has been developed by Fukui for the full potential energy surfaces[37,38] and has been further elaborated by many authors since.[26,75,80–85] In simple terms IRC represents the center curve of the reaction path region that follows the invariant energy gradient.

In particular, in Cartesian coordinates the IRC curve satisfies the following simple condition

$$\nabla_{\perp}U(\overline{X}) = \nabla U(\overline{X}) - \vec{n}(\overline{X})\frac{\vec{n}(\overline{X})\cdot \nabla U(\overline{X})}{\vec{n}(\overline{X})\cdot\vec{n}(\overline{X})} = \vec{0} \quad (6)$$

where $\vec{n}(\overline{X})$ is the curve tangent and $\vec{0}$ is the null vector.

Importantly, for nonlinear coordinates the direction of the gradient vector has to be corrected using the corresponding contravariant metric tensor G that potentially depends on all $3N$ degrees of freedom:

$$G = (g^{ij}) = \left(\sum_{k=1}^{3N}\frac{\partial q_i}{\partial x_k}\frac{\partial q_j}{\partial x_k}\right) \quad (7)$$

otherwise different nonlinear coordinate systems will yield different reaction paths for the same stationary points.[37,38,80–83,85] Thus, to be invariant the transition path curve in nonlinear coordinates must satisfy the following more complicated condition

$$(G\nabla U(\overline{Q}))_{\perp} = G\nabla U(\overline{Q}) - \vec{n}(\overline{Q})\frac{\vec{n}(\overline{Q})\cdot(G\nabla U(\overline{Q}))}{\vec{n}(\overline{Q})\cdot\vec{n}(\overline{Q})} = \vec{0} \quad (8)$$

where $\vec{n}(\overline{Q})$ is the curve tangent in the nonlinear coordinates.

Both eqs 6 and 8 apply also to the adiabatic energy surfaces. Because the system of eq 8 is somewhat complicated by the need to compute the metric tensor, the ggaHFB method employs Cartesian coordinates for the path curve optimization instead of the generalized coordinates.

5. Minimum Free Energy Transition Path Ensemble. Using the free energy gradient, the ggaHFB method generalizes the concept of the Fukui's IRC[37,38] to free energy surfaces. In deriving the free energy gradient the SCS degrees of freedom orthogonal to RCS are averaged over, which results in each point in the RCS representing an ensemble. Thus, the ggaHFB method finds continuous curves that connect the provided reactant and product ensembles through a series of transition and intermediate state ensembles. These curves must satisfy the condition that the invariant free energy gradient be tangential to the path curve at any point. In particular, the ggaHFB method uses the straightforward generalization of eq 6 to free energy surfaces in Cartesian coordinates:

$$\nabla_{\perp}W^{u}(\langle\overline{x}\rangle^{b}) = \nabla W^{u}(\langle\overline{x}\rangle^{b}) - \vec{n}(\overline{x})\frac{\vec{n}(\overline{x})\cdot\nabla W^{u}(\langle\overline{x}\rangle^{b})}{\vec{n}(\overline{x})\cdot\vec{n}(\overline{x})} = \vec{0} \quad (9)$$

As noted above, working with nonlinear coordinates requires computing logarithmic Jacobian corrections to the free energy gradient. Furthermore, finding invariant paths requires additional metric tensor corrections.[19,74] No such complications arise in Cartesian coordinates, which is why the ggaHFB method employs these coordinates to optimize transition path ensembles.

6. Transition Path Optimization. To optimize a transition path in Cartesian coordinates, we take K unique configurations $\{\overline{Q}_k\}_{k=\overline{1,K}}$ that gradually progress from the reactant to the product and assign them to a uniform grid $\{\alpha_k = (k-1)(K-1)\}_{k=\overline{1,K}}$ with mesh size of $\Delta\alpha = 1/(K-1)$. If initial configurations $\overline{Q}_k = \overline{Q}(\alpha_k)$ are unavailable they could be derived via a linear interpolation or by the activated evolution procedure[21] that is similar to the growing string method.[86] Using these K configurations, we obtain up to K corresponding Fourier amplitudes for each degree of freedom by applying the standard Fourier transform integration with the trapezoidal rule on the grid[87]

$$b_n^i = \sum_{k=1}^{K-1}(f_n^{i,k} + f_n^{i,k+1})\Delta\alpha \quad (10)$$

where $f_n^{i,k} = [q_i(\alpha_k)–q_i(0)–(q_i(1)–q_i(0))\alpha_k]\sin(n\pi\alpha_k)$.

This procedure globally interpolates between all the K points, yielding a continuous Fourier curve[21,88] which is an analytical function of a progress variable $\alpha \in [0;1]$:

$$q_i(\alpha) = q_i(0) + (q_i(1) - q_i(0))\alpha + \sum_{n=1}^{K} b_n^i\sin(n\pi\alpha) \quad (11)$$

We then redistribute the K beads along the path curve such that they conform to a particular metric. Usually, we reposition the beads to make the arc lengths between adjacent beads of equal length in the RCS.

The newly redistributed beads serve as reference beads to compute the corresponding adiabatic potential energy gradients or the free energy gradients via the evolution procedures described in sections 2 and 3. Thus, for each reference bead $\bar{q}_k^{ref} = \bar{q}(\alpha_k^{ref})$, the evolution returns either the minimized $[\bar{q}]_k^b = ([q_1]_k^b, \cdots, [q_S]_k^b)$ or the averaged bead $\langle\bar{q}\rangle_k^b = (\langle q_1\rangle_k^b, \cdots, \langle q_S\rangle_k^b)$ also called the “raw evolved bead”.

The ggaHFB method borrows the idea of redistributing beads along the curve and reparametrizing the curve given the redistributed beads from the string method.[73–75,86] All the other essential ingredients of the ggaHFB method, such as the multidimensional energy gradient derived on the fly from the harmonic biasing potential and the Fourier representation of both the path and of the corresponding energy gradient (see below) employed in the energy profile integration via generalized line integral as well as optimization strategies, have been obtained from sources independent of the string method despite apparent similarity.[21–23,37–40,79,88,104]

In the following discussion, we omit the complementary SCS coordinates for clarity. These coordinates are assumed to be either completely minimized or averaged over and do not explicitly affect either the path or its energy. Optimization implies that the SCS coordinates are passed along either through dynamics restart files or through the complete coordinate files. In addition, it is assumed that the changes in the SCS coordinates between the beads are continuous.

For brevity, we only discuss how to drive optimization of the transition path ensembles and compute the corresponding free energy profiles. The same strategies apply to finding the transition paths on adiabatic potential energy surfaces and computing the corresponding energy profiles. In this case, the adiabatic potential energies could also be calculated exactly for all the points along the path and compared to those computed using the ggaHFB's generalized line integral formalism.

Substituting the raw evolved beads into eq 3a gives estimates of the free energy gradients for each bead. These gradients are then used in the steepest descent step to generate the "enhanced evolved beads":

$$\overline{q}_k^{SD} = \langle \overline{q} \rangle_k^b + \gamma_k \nabla W^u(\langle \overline{q} \rangle_k^b) \qquad (12)$$

Here γ_k is the parameter that controls the SD step size for the k^{th} bead. In the present paper we use the uniform step size parameter γ for all the beads for simplicity.

Following the Fourier transform of the enhanced beads to obtain new Fourier amplitudes, redistribution of the beads along the resulting curve provides new reference beads. These reference beads are realigned to maintain the coordinate system. For this purpose we invoke a mass-weighted best-fit procedure in a suitable space, usually RCS, to enforce the Eckart conditions on the beads.[22,79,89–91] In cases where only a few coordinates are available for the best fit or if their geometric arrangement breaks down the standard best fit procedure, simpler alignment methods could be used. The final realigned beads then replace the previous reference beads in the next round of evolution. This procedure is repeated until convergence of the path, i.e. until the path curve changes cease. The final optimized curve represents an invariant minimum free energy transition path ensemble that satisfies the Fukui's IRC criteria.[37,38]

The convergence rate of the ggaHFB method depends to some extent on the employed bias force constant and step size parameter. Therefore, devising an optimization strategy to achieve the fastest convergence possible is desirable and is an active area of research in our laboratory.

7. Computing the Free Energy Along the Fourier Path. Given a Fourier path in the generalized multidimensional coordinate space and the corresponding free energy gradients, we can compute the free energy profile along that path via the generalized line integral formalism. To achieve the highest accuracy, we Fourier transform both the evolved beads (4) and the corresponding free energy gradients (5) along the path. With the continuous Fourier representations of the forces and the path, we could then analytically evaluate the corresponding reversible work line integral passing through the evolved beads:

$$W^u(\alpha) = \sum_{i=1}^{S} \int_0^{\alpha} \left[\frac{\partial W^u(\alpha)}{\partial q_i} q'_i(\alpha) \right] d\alpha \qquad (13)$$

In practice, we evaluate the generalized line integral of the second order in eq 13 on a fine uniform grid with $L \gg K$ quadrature points.

This procedure provides the free energy or the potential of mean force (PMF) profile as an analytical function of the progress variable. Unlike umbrella sampling with WHAM, the interpolation-based ggaHFB free energy integration procedure does not require overlap between the windows. Furthermore, the ggaHFB integration procedure allows natural decomposition of the free energy into contributions from individual coordinates.

The analytical form of the energy profile and that of the corresponding path provided by the ggaHFB method renders pinpointing the energy extrema and their accurate RCS coordinates particularly trivial. One can easily find the values of the progress variable α corresponding to extrema on the energy profile and then substitute these values into eq 11 to get the matching structures.

8. Summary of the ggaHFB Methodology. In summary, the ggaHFB method finds the Fukui's IRC curves on the adiabatic potential energy surfaces and further generalizes this approach to Cartesian free energy surfaces. Thus, the ggaHFB method finds either minimum adiabatic potential energy paths or minimum free energy transition path ensembles via a gradient driven optimization procedure. Optimizing the transition paths and path ensembles in Cartesian coordinates bypasses the need to calculate the corresponding metric tensors. The optimized transition paths provide structural and energetic information about all the intermediates and transition states connecting given reactants and products at once. Furthermore, the global Fourier representation of the path and the forces provide useful means to control various aspects of the path optimization and ultimately makes the ggaHFB optimization extremely robust.

Independent from the path optimization, the ggaHFB method is a practical alternative to the conventional approach to free energy calculations via umbrella sampling with WHAM. Advantageously, the ggaHFB method is histogram-free, which makes it applicable to cases with arbitrary many dimensions. Even though ggaHFB uses somewhat stiff springs, it does not require the overlap between the windows to integrate the free energy profile. Additionally, the Cartesian version of the ggaHFB method avoids the need to compute the logarithmic Jacobian correction that is required if either WHAM or ggaHFB is used with nonlinear coordinates such as bond distances, angles, dihedrals, etc. to compute free energy profiles.[23,92] Finally, the energy profiles can be straightforwardly decomposed into contributions from the individual degrees of freedom that could be useful for analysis and design purposes.

Results

1. Minimum Free Energy Transition Path Ensembles. To explore the free energy of the α_L to pII backbone transition of a TATTVGYG signature peptide in water and the effect of the V76 side chain rotation, we use the ggaHFB method with two reactive coordinate spaces (RCSs) of different dimensionalities. Specifically, we include all heavy atoms of the peptide into RCS1 and derive RCS2 from RCS1 by excluding side chain atoms. The RCS1 surface provides the free energy of the backbone configuration subject to a particular side chain orientation. In contrast, the RCS2 surface provides the free energy of the backbone configuration irrespective of the side chains. Unless otherwise stated, throughout this work we employ molecular dynamics in the isothermal isobaric NPT ensemble at 298 K and 1 atm using the CHARMM22 molecular mechanical force field[59,60] with the CHARMM-modified TIP3P explicit water model[93–97] to derive all the required free energy gradients.

Our preliminary free-energy optimization runs revealed that both the α_L and pII states at position 77 are local free-energy minima of the isolated peptide in water. Interestingly, the partially flipped, nonconducting conformation observed by X-ray crystallography at low K^+ concentration (PDB code 1R3K) is unstable by itself in water despite the rotation of

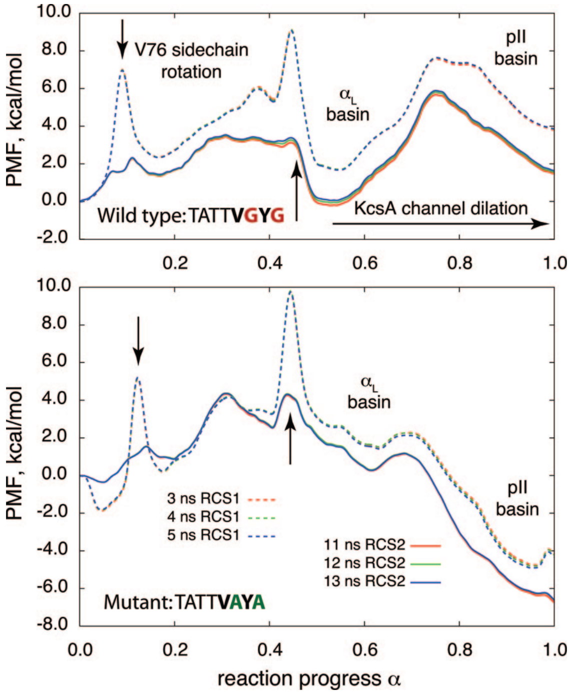


Figure 1. Cumulative PMFs for the conformational transition of the signature TATTVGYG KcsA peptide and its AYA mutant from the α_L to pII state in explicit water on RCS1 and RCS2 free-energy surfaces at different collection times. The ggaHFB method employed 89 beads to integrate the free energy profile. Arrows point to the V76 side chain rotations.

the V76 side chain away from the high K^+ concentration, conducting conformation (PDB code 1R3J). During optimization of the peptide from the partially flipped state its backbone, but not the V76 side chain, collapses to the conducting state conformation.

Therefore, to study the full range of the peptide flip, we have constructed an initial path that includes both α_L and pII states of G77 backbone. To assess the effect of the V76 side chain rotation, we have included two such rotations by requiring that the end points have the same V76 side chain orientation, matching that of the conducting state. Furthermore, we have inserted the crystallographic nonconducting state with partially flipped backbone and rotated V76 in the middle of the path (refer to the Supporting Information for details).

Performing thorough optimization on the RCS1 free energy surface (see the Supporting Information), we obtain an intrinsic transition path ensemble for the G77 backbone conformational transition from α_L to pII state in the TATTVGYG peptide in water. Using the RCS1-optimized path ensemble as the reference we then compute the final free-energy profile, both coupled with (RCS1) and uncoupled from (RCS2) the V76 side chain rotation. To elaborate on the energetics of the backbone transition further, we compute an analogous optimal path and the free-energy information for a GYG to AYA mutant. The resulting cumulative PMFs at different collection times are depicted in Figure 1, and the representative structures of the wild type peptide are shown in Figure 2.

The PMFs at the RCS1 level display two events involving the V76 rotation as two sharp peaks with barrier heights ranging from 5 to 8 kcal/mol in both directions. Interestingly, the V76 side chain rotation from the conducting into nonconducting orientation destabilizes the α_L state by 2 to 3 kcal/mol, indicating that the hydrophobic interaction between the V76 and Y78 side chains provides additional stabilization of the α_L state in the conducting conformation. The α_L to pII backbone transition at position 77 follows the second, restoring V76 side chain rotation. In the wild type, the free energy barrier for the backbone transition given a specific orientation of the side chains (the RCS1 PMF) has a forward barrier of 6.0 kcal/mol. The pII state is 2.2 kcal/mol less favorable than the α_L state and converts back with a barrier of 3.8 kcal/mol. In sharp contrast, in the mutant the forward barrier is only 0.7 kcal/mol and the pII state is 6.3 kcal/mol more favorable than the α-strand. Restoring the α-strand[52–54] state in the mutant requires surmounting a high 7.0 kcal/mol free-energy barrier.

Setting the side chains free (the RCS2 PMF) permits evaluating the free energy of the backbone transition alone. As is seen from Figure 1, switching to RCS2 space collapses the sharp peaks (labeled with arrows) corresponding to the V76 side chain rotations but leaves the portion of the PMF underlying the backbone transition from the α_L to the pII state virtually unchanged. In particular, for the wild type peptide the forward activation barrier is 5.9 kcal/mol, and the pII state is still less stable than the α_L by slightly smaller 1.7 kcal/mol. Restoring the conducting state requires overcoming a slightly higher barrier of 4.2 kcal/mol. In contrast, the mutant exhibits a forward barrier of 0.9 kcal/mol and the relative pII state stabilization energy of 7.0 kcal/mol that makes the reverse barrier increase to 7.9 kcal/mol.

2. Minimum Adiabatic Potential Energy Paths. To explicitly evaluate the effect water has on the conformational transitions of the signature peptide and to further demonstrate the capabilities of the ggaHFB method, we have computed the minimum adiabatic potential energy paths for the wild type TATTVGYG peptide and its GYG to AYA mutant in gas phase. Both peptides have three threonine and one tyrosine residues with rotatable OH bonds that were averaged over in the minimum free energy transition path ensembles computed in water. The orientation of these hydrogens significantly perturbs the overall potential energy; therefore, we include these four hydrogen atoms in the reactive coordinate space. Thus, by adding the polar hydrogen atoms to the all-heavy-atom RCS1 we derive the RCS1h for adiabatic potential energy path optimization.

We have to assign some initial values to the tyrosine and threonine OH groups, which have two and three rotameric states, respectively, in order to compute the adiabatic potential energy paths. The total number of possible initial path configurations is therefore $2^1 \times 3^3 = 54$. To control the configurations, we follow the Protein Data Bank atom naming convention and use dihedral angles Cε1-Cζ-Oη-Hη and Cα-Cβ-Oγ-Hγ for the tyrosine and the threonines, respectively. Here, we arbitrarily choose dihedral angles of 180, 180, −30, and 0 degrees for the T72, T74, T75, and Y78, respectively, as the initial conditions for the path

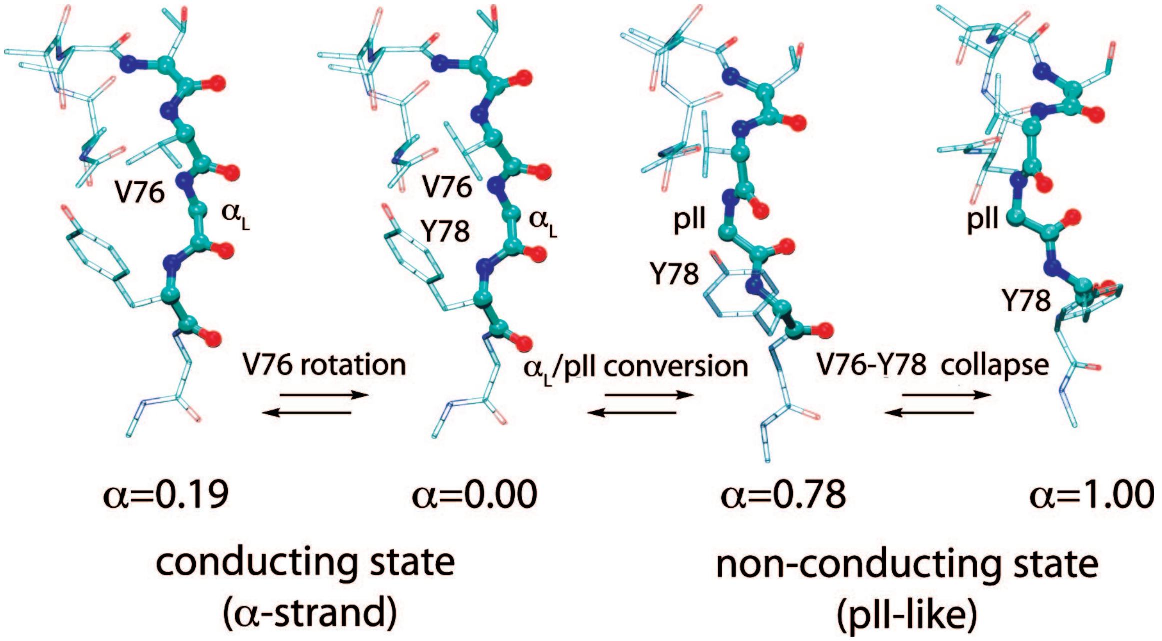


Figure 2. Representative structures from the free energy transition path ensemble of the wild type TATTVGYG signature peptide of the KcsA selectivity filter in explicit water. The values of the progress variable α provided relate structures to the free energy profile of the wild type peptide in Figure 1.

optimization. To prepare the initial path with these conditions, we fix the RCS1 coordinates and apply stiff harmonic restraints of 1000 kcal/(mol·rad^2) on the corresponding dihedral angles during an optimization of the hydrogen positions.

Because the optimization on adiabatic potential energy surfaces with the bare CHARMM22 molecular mechanical force field is relatively inexpensive, we have initiated the ggaHFB optimization using 89 beads. Using 89 beads is sufficient to integrate the adiabatic potential energy, given the initial orientation of the four OH bonds. Nevertheless, optimization of the OH groups requires increasing the number of beads further to correctly integrate the adiabatic potential energy. The increase reflects the fact that rotations of the OH groups correspond to small changes in the RCS1h, resulting in very sharp transitions along the path. Although proper integration could be achieved by locally increasing the number of beads at the sharp transitions leading to nonuniform bead distributions,[23] in the present work we use uniform grid for simplicity. Thorough path optimization increases the overall path length dramatically, in the end requiring 705 beads to properly integrate the adiabatic potential energy along the path.

The final paths in the gas phase have little if any resemblance with the paths optimized in water and exhibit a greater number of local minima and transition states. For the wild type peptide, the α-strand[52–54] disappears almost completely. First, the G79 residue spontaneously flips into the C_5 conformation and then converts into the C_{7ax} conformation. In the flipped configuration on the reactant side of the path G79 forms a hydrogen bond with the OH group of T71 using its carbonyl oxygen. The G79 residue flip significantly perturbs the rest of the α-strand, which quickly collapses further residue-by-residue along the path.

In the mutant, the α-strand is annihilated completely in the reactant basin, where the A79 along with the A77 residues flip into the C_{7ax} conformation. The four residues V76, A77, Y78, and A79 surround the T71 residue like a belt, with alternating axial and equatorial configurations, namely C_{7ax}, C_{7eq}, C_{7ax}, and C_{7eq}, respectively. During the optimization the mutant pathway deviates substantially from that of the wild type.

Figure 3 depicts the corresponding adiabatic energy profiles that underscore the complexity of the changes in the gas phase. It also provides the benchmarks for the adiabatic potential energy integration via the generalized line integral formalism. In particular, comparison of the line integral energies with the exact adiabatic potential energies from the CHARMM22 force field shows the accumulated errors of 0.07 and 0.12 kcal/mol for the wild type and mutant adiabatic energy profiles, respectively. We consider this a very good agreement between the generalized line integral energy and the exact energy profiles.

The V76 side chain rotations (labeled with arrows) have been preserved in both wild type and mutant paths, although in some cases they have been coupled with other structural rearrangement as seen in Figure 3. The forward and reverse barrier heights for the V76 side chain rotation vary but are similar to those in water.

Overall the gas phase structures are more compact than the ones in water and establish as many intramolecular hydrogen bonds as possible. Given the complexity of the adiabatic paths and their divergence from the structures obtained by either the X-ray crystallography or by the free energy optimization in water, we omit a detailed description of the structural changes along the path and simply provide the corresponding trajectories in Supporting Information.

3. Comparison of the MM and QM Energy Profiles.

A. Gas Phase. Because the present paper investigates an

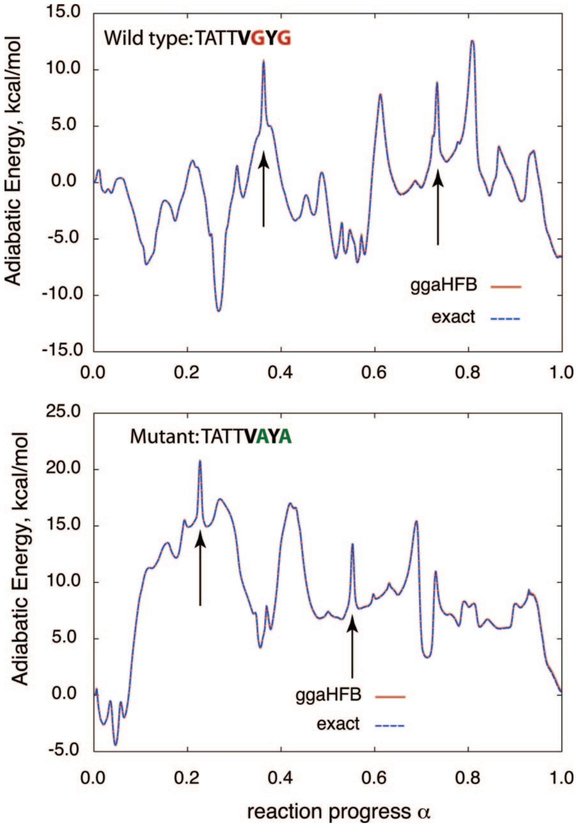


Figure 3. Adiabatic potential energy profiles along the optimized reaction paths of the signature TATTVGYG KcsA peptide and its AYA mutant in gas phase. The ggaHFB method employed 705 beads to integrate the potential energy with the RCS1h (see text for details) - solid lines; the CHARMM22 exact energies are shown in dashed lines. Arrows point to the rotations of the V76 side chain.

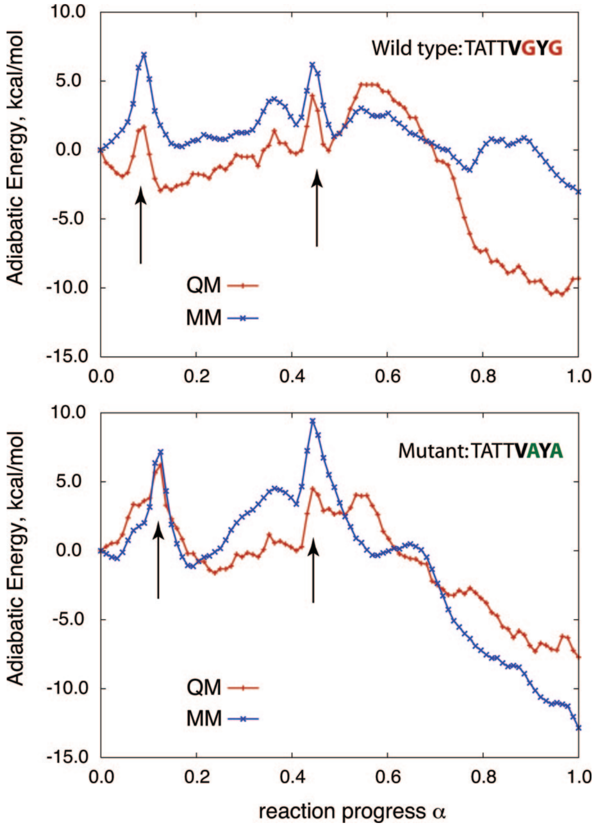


Figure 4. Gas phase single point energy profile along the α-strand to pII state conformational transition of the signature TATTVGYG KcsA peptide and its AYA mutant in water. MM - uses the bare CHARMM22 force field, QM - B3LYP/6−31G(d) Density Functional Theory model. Arrows point to the rotations of the V76 side chain.

important conformational transition of the TATTVGYG signature peptide from the KcsA potassium channel, it might be useful to assess the molecular mechanical (MM) force field employed. Of particular interest is evaluating the energetics of the signature peptide and its mutant along the path optimized in water. To establish useful benchmarks, we first compute the gas phase adiabatic energy profiles along the minimum free energy transition path ensembles in the RCS1. In particular, we compare the MM energy profiles with one of the most popular density functional theory models, namely B3LYP, with the 6−31G(d) basis set as a high-level quantum mechanical (QM) model (see the Supporting Information for details). This model is not expected to produce accurate energy profiles when it comes to dispersion interactions between the Y78 residue and the V76 side chains and hence should be used with caution.[98–101]

Figure 4 shows the corresponding adiabatic potential energy profiles for the peptides with the rotatable OH bonds fixed at the conformation used as initial condition for the path reoptimization in gas phase. Interestingly, the energy profiles obtained with MM and the QM models for the same path differ substantially. The V76 side chain rotation barriers appear reduced in the QM model.

Both the MM and QM models favor the pII state over the α-strand. The QM model predicts the α-strand to be much less stable than the pII state in the wild type but relatively more stable in the mutant peptide. In contrast, the MM model suggests that the α-strand is much more stable in the wild type peptide, not the mutant. It is likely that the adiabatic energy surfaces of the MM and QM models are significantly different in the gas phase, and such single point energy profile comparisons should be taken with caution.

B. Implicit Solvent. As mentioned above, we are primarily interested in the energetics of the peptides in water and not in the gas phase. After all, the transition path ensembles for the functional conformational transitions of the peptides have been optimized using the MM model in the explicit water. To compare the MM with QM models in water, we choose the Generalized Born Molecular Volume (GBMV)[47,64] and Polarizable Continuum Model (PCM)[65–68] implemented in Gaussian 03[102] implicit solvent models, respectively. Using the free-energy optimized transition path ensembles in explicit water as reference profiles, we have performed optimization of all the degrees of freedom orthogonal to the RCS1 and subject to the same dihedral restraints on the rotatable OH bonds as discussed above with the MM-GBMV model and then computed single point QM-PCM energies for all the beads along the path (see the Supporting Information for details). The results are provided in Figure 5.

The MM-GBMV model yields the free-energy profiles that are in very good agreement with the explicit solvent

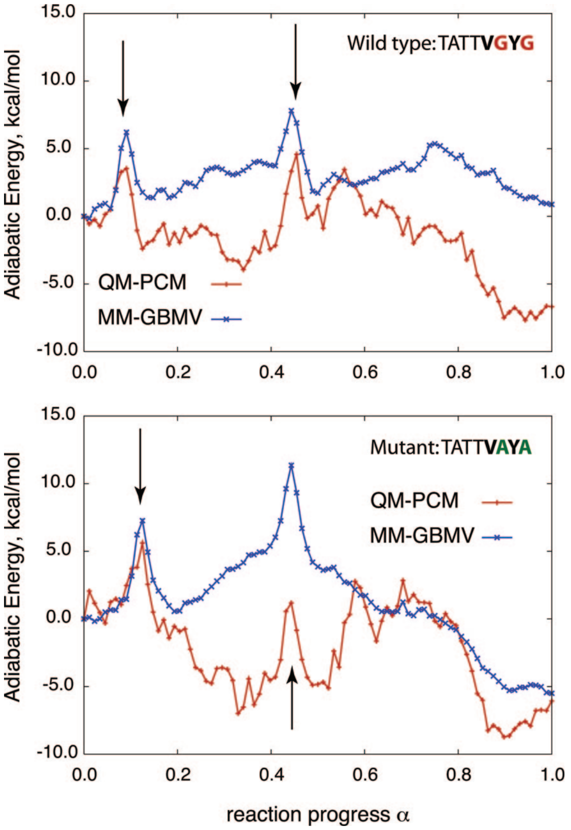


Figure 5. Implicit solvent single point energy profile along the α-strand to pII state conformational transition of the signature TATTVGYG KcsA peptide and its AYA mutant in water. MM-GBMV - uses the CHARMM22 force field with the Generalized Born Molecular Volume implicit solvent model, QM-PCM - B3LYP/6−31G(d) Density Functional Theory with the Polarizable Continuum Model implicit solvent model. Arrows point to the rotations of the V76 side chain.

calculations (Figure 1), thus further validating the intrinsic free energy profiles. We cannot expect better agreement between the two profiles given the fixed conformations of the rotatable OH bonds necessary to compute the adiabatic energy profile with GBMV.

The QM-PCM model produces energy profiles very different from those of the MM-GBMV model, and most importantly does not favor the α_L state over the pII state in the wild type peptide. In the mutant, on the other hand, the α_L state remains unstable with the QM-PCM model.

Because we have performed optimization on the water modified RCS1 free energy surfaces of the peptides with the MM model and explicit water any comparison with other surfaces that have not been optimized do not warrant good agreement, unless the surfaces are exactly the same. This conflict could in principle be resolved by the QM-PCM optimization of the product, reactant, and a few key intermediates, which unfortunately presents a significant challenge at present.

Discussion

The finding that the signature peptide taken from the partially flipped nonconducting state (PDB code 1R3K) collapses back into the conducting state in water despite the V76 side chain rotation shows the intrinsic width of the peptide α_L basin. Furthermore, it suggests that either the channel provides additional interactions to stabilize the partially flipped backbone structure or that only one of the four strands of the tetrameric channel undergoes the full transition. If the latter symmetry breaking were to occur, the apparent configuration observed by X-ray crystallography would correspond to the average over four strands, thus artificially reducing the extent of the transition in a single peptide.

The fact that the free energy profiles for the α_L to the pII transition are relatively insensitive to the position of the side chains (see Figure 1) reflects the robustness of the backbone transition. The forward and reverse free energy activation barriers in the wild type peptide are 5.9 and 4.2 kcal/mol. Interestingly, previous calculations of an even lower dimensional PMF for a similar transition inside the wild type KcsA channel in the presence of two ions gave a rough estimate of the free energy barrier between 0.5 and 4.0 kcal/mol.[8] In sharp contrast, the mutant exhibits a forward barrier of 0.9 kcal/mol and the reverse barrier of 7.9 kcal/mol.

The width of the α_L basin in the intrinsic free energy profile of a single peptide might determine the range of a local dilation/contraction of the tetrameric KcsA channel at the V76 carbonyl ring. If the V76 carbonyl were pushed away from the channel axis beyond the limits of the α_L basin, the peptide would go over the transition barrier and into the nonconducting pII state. We emphasize that by local dilation/contraction of the channel we imply the change in the distance between the V76 carbonyls associated with the backbone motions within the bounds of the α_L basin and not with the transition from α_L to pII or back.

The full α_L to pII transition has been demonstrated to be unnecessary for the ion selectivity, at least in a synthetic channel with the D-Ala residue in place of G77.[3] Note that the wild type KcsA channel, in addition to K^+, permits ions of larger size, namely Cs^+ and Rb^+, which are expected to pass the V76 carbonyl ring without triggering the transition from α_L to pII.[4] Such a wide range of dilation/contraction would not have been possible if G77 was substituted for regular Ala, as the width and the depth of the α_L basin would have been dramatically reduced as seen from the PMFs for the mutant depicted in Figure 1. Note however, that the AYA mutant would be sterically prevented from forming the conducting α-strand conformation in the tetrameric channel.[3]

In an effort to validate the results obtained with the MM force field in explicit water we have profiled the energetics along the paths using MM and QM methods both in gas phase and in implicit solvent. The results of these calculations are summarized in Figures 4 and 5 that highlight the stark disagreement between the MM and QM models. Although QM models usually have higher fidelity than MM models, the particular DFT method used in this work, namely the B3LYP functional, is well-known to fail to account for short-range dispersion interactions necessary to properly describe the energetics of the hydrophobic interactions such as those between V76 and Y78 side chains.[98−101] Higher level, more expensive *ab initio* methods that properly account for the dispersion suggest that the interactions between the CH bonds of the V76 and the phenol ring of the Y78 could favor the α-strand by about 1 kcal/mol.[101,103] Additional discrepancies between the MM and QM in this work may arise due to the fact that no optimization has been performed at the QM level of theory. Therefore, the differences between the QM and MM models should be interpreted with caution.

It appears that the stability of the pII state in the wild type peptide might be overestimated by the QM model with implicit water, because it would require at least 20 kcal/mol to assume the conducting conformation in the tetrameric channel. On the other hand the MM model with implicit water predicts the α_L and pII configurations to be nearly degenerate. If the QM-PCM model more accurately reproduced the energetics of the solvated peptide even without optimization, the ground or resting state of the wild type KcsA channel would be the nonconducting pII state. Thus the channel would have to be activated by a conformational change from the pII resting state into the α-strand state to conduct ions. This would only be possible due to a strong perturbation such as strong attraction of the ions in the lumen of the tetrameric channel to its carbonyl oxygens.

The switching between the nonconducting and conducting state and the functional contraction/dilation of the tetrameric KcsA channel would require a certain balance between the electrostatic repulsion of the V76 carbonyls and the free energy of the backbone rotation of the residue at position 77. Because the electrostatic repulsion can be relaxed by transiently flipping one or more of the carbonyls out from the α_L into the pII state, the filter must also ensure to favor the α_L over the pII state at least in the presence of ions in the lumen of the filter. The potassium channel seems to have achieved the α-strand stabilization by using a G residue that has a high propensity for the α_L configuration at position 77 and in addition by the hydrophobic interaction between V76 and Y78 side chains. The importance of the hydrophobic interaction is supported by the experimental observation that the V76A mutant abrogates tetrameric assembly of the channel.[58] Taking the above into consideration, it appears that the free energy profiles computed with the MM model in explicit water agree better with the proposal than the corresponding single point energy profiles obtained with B3LYP/6−31G(d)-PCM QM model.

In the absence of the actual tetrameric channel in our model, the bulk water better reproduces the environment of the KcsA selectivity filter than the gas phase and therefore provides useful insights into the channel function. In particular, the differences between the adiabatic energy maps of the peptide residues in water and gas phase suggest that the gas phase transition pathways must deviate strongly from those of the transition path ensembles optimized in water. This is particularly true of the α_L region that is forbidden in gas phase.[48,49] The ggaHFB optimization of the adiabatic paths in gas phase explicitly demonstrates that water plays an active and important role in defining the intrinsic path and the energetics of the peptide backbone transition.

The outcome of the gas phase optimization can be predicted based on the previous studies of the glycine and alanine dipeptides.[48,49] In particular, the referred work demonstrated that the α_L configurations collapse into the C_{7ax}, while pII configurations collapse into the C_{7eq}.[48,49] These are the exact changes we observe upon the adiabatic potential energy optimization in gas phase. The final optimized paths in gas phase are rather complex (see Figure 3) and seem irrelevant for the functional transition of the selectivity peptide in the KcsA channel. On the other hand the pathways in water show very good qualitative agreement with the peptide conformations observed in the tetrameric channel.

Finally, based on the present findings, we are able to propose a novel hypothesis for the mechanism of ion selectivity in the tetrameric KcsA channel. Specifically, we conjecture that in its conducting α-strand state the carbonyl rings should contract around the ion entering the channel and that this contraction would propagate to the nearby carbonyl rings along the channel axis (see Figure 1S in the Supporting Information for an illustration). Because ions are believed to pass the KcsA filter stripped of all but two water molecules that cotranslate with the ion while hydrogen bonding to the carbonyls, these water molecules will experience greater difficulty to pass neighboring carbonyl rings due to the contraction, in turn impeding the ion movements along the channel.

This hypothesis could explain why the channel selects larger K^+ over smaller Na^+ ions. Specifically, we anticipate that smaller Na^+ ions would contract the carbonyl rings to a greater extent than the larger K^+ ions thus impeding the passage of the cotranslating water molecules to a greater degree. With water passage impeded the ions themselves must in turn slow down.

Our hypothesis suggests that in the absence of ions the KcsA channel should stay open to water permeation unless one of the four V76 carbonyls flips out pinch-shutting the channel. Indeed, the KcsA channel has been experimentally demonstrated to conduct water in the absence of permeating ions.[51] The partial flipping of the carbonyls (while still within the α_L basin) might serve a selectivity purpose, whereas a complete flip (transition into the pII basin) can be used to gate the channel.[8] To provide further support of this hypothesis we are currently performing an optimization of several transition path ensembles for ion−water copermeation through the tetrameric KcsA selectivity filter. The results of this work will be reported in a forthcoming publication.

To conclude, we have explored an intrinsic free energy landscape of an important functional transition of the signature peptide from the KcsA selectivity filter that is responsible for locally dilating/contracting the channel at the V76 carbonyl ring, in addition to switching the channel between conducting and nonconducting states. We have found that the wild type peptide intrinsically favors the conducting state due to the combination of the high G77 backbone propensity for the α_L configuration and the stabilizing hydrophobic interaction between V76 and Y78 side chains. In sharp contrast, the mutant strongly favors the nonconducting state. However, additional steric effects in the tetrameric channel that are absent in the present study are expected to prevent formation of the conducting conformation in the mutant.

We have found the α_L to pII transition to be exceptionally robust and intrinsically funneled toward the conducting state in the wild type KcsA peptide at the MM level with explicit water. Although the intrinsic free energy profiles have been validated using the MM with an implicit water model, efforts to gauge performance of the MM model against the QM model indicated that our results should be interpreted with caution. Based on the QM-PCM model it may be possible that the ground state of the channel in the absence of ions could in fact be the nonconducting state and that the conducting state would only form upon ion entrance into the lumen of the channel. Nevertheless, the present study has allowed us to propose a novel hypothesis for the ion selectivity within the KcsA channel in which local contraction of the channel interior in response to the ion presence regulates copermeation of water through the channel to a degree that is inversly proportional to the ion size. Work is currently underway in our laboratory to test the proposed hypothesis in the tetrameric KcsA channel model. We hope that the present work will stimulate future transition path ensemble studies of rare events in complex molecular systems.

Acknowledgment. This work was in part supported by grants from NIH, NSF, NBCR, CTBP, and HHMI. Computer time on clusters at CTBP, NBCR, and the Department of Bio-Engineering at UCSD is greatly appreciated. M.F. was supported in part by an NIH Molecular Biophysics Training Grant: GM-08326, and subsequently by an NSF Graduate Research Fellowship.

Supporting Information Available: Peptide coordinates from the optimized transition path ensembles together with details of the transition path optimization and PMF calculations, a cartoon illustration of the ion selectivity hypothesis, and details of the B3LYP/6−31G(d)-PCM calculations. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Zhou, Y.; MacKinnon, R. The occupancy of ions in the K+ selectivity filter: charge balance and coupling of ion binding to a protein conformational change underlie high conduction rates. *J. Mol. Biol.* **2003**, *333* (5), 965–975.

(2) Zhou, Y.; Morais-Cabral, J. H.; Kaufman, A.; MacKinnon, R. Chemistry of ion coordination and hydration revealed by a K+ channel-Fab complex at 2.0 Å resolution. *Nature* **2001**, *414*, 43–48.

(3) Valiyaveetil, F. I.; Leonetti, M.; Muir, T. W.; MacKinnon, R. Ion Selectivity in a Semisynthetic K+ Channel Locked in the Conductive Conformation. *Science* **2006**, *314* (5801), 1004–1007.

(4) Lockless, S. W.; Zhou, M.; MacKinnon, R. Structural and Thermodynamic Properties of Selective Ion Binding in a K+ Channel. *PLoS Biol.* **2007**, *5* (5), 1079–1088.

(5) Berneche, S.; Roux, B. Molecular dynamics of the KcsA K+ channel in a bilayer membrane. *Biophys. J.* **2000**, *78* (6), 2900–2917.

(6) Aqvist, J.; Luzhkov, V. Ion permeation mechanism of the potassium channel. *Nature* **2000**, *404*, 881–884.

(7) Berneche, S.; Roux, B. Energetics of ion conduction through the K+ channel. *Nature* **2001**, *414*, 73–77.

(8) Berneche, S.; Roux, B. A gate in the selectivity filter of potassium channels. *Structure* **2005**, *13*, 591–600.

(9) Mashl, R. J.; Tang, Y.; Schnitzer, J.; Jacobsson, E. Hierarchical Approach to Predicting Permeation in Ion Channels. *Biophys. J.* **2001**, *81* (5), 2473–2483.

(10) Guidoni, L.; Carloni, P. Potassium permeation through the KcsA channel: a density functional study. *Biochim. Biophys. Acta* **2002**, *1563* (1−2), 1–6.

(11) Compoint, M.; Boiteux, C.; Huetz, P.; Ramseyer, C.; Girardet, C. Role of water molecules in the KcsA protein channel by molecular dynamics calculations. *Phys. Chem. Chem. Phys.* **2005**, *7* (24), 4138–4145.

(12) de Haan, H. W.; Tolokh, I. S.; Gray, C. G. Nonequilibrium molecular dynamics calculation of the conductance of the KcsA potassium ion channel. *Phys. Rev. E* **2006**, *74* (3), 030905.

(13) Grabe, M.; Bichet, D.; Qian, X.; Jan, Y. N.; Jan, L. Y. K+ channel selectivity depends on kinetic as well as thermodynamic factors. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103* (39), 14361–14366.

(14) Kraszewski, S.; Boiteux, C.; Langner, M.; Ramseyer, C. Insights into the origins of the barrier-less knock-on conduction in the KcsA channel: molecular dynamics simulations and ab initio calculations. *Phys. Chem. Chem. Phys.* **2007**, *9* (10), 1219–1225.

(15) Domene, C.; Vemparala, S.; Furini, S.; Sharp, K.; Klein, M. L. The role of conformation in ion permeation in a K+ channel. *J. Am. Chem. Soc.* **2008**, *130* (11), 3389–3398.

(16) Gwan, J.-F.; Baumgaertner, A. Cooperative transport in a potassium ion channel. *J. Chem. Phys.* **2007**, *127* (4), 045103:1–10.

(17) Noskov, S. Y.; Berneche, S.; Roux, B. Control of ion selectivity in potassium channels by electrostatic and dynamic properties of carbonyl ligands. *Nature* **2004**, *431*, 830–834.

(18) Noskov, S. Y.; Roux, B. Ion selectivity in potassium channels. *Biophys. Chem.* **2006**, *124* (3), 279–291.

(19) Maragliano, L.; Vanden-Eijnden, E. On-the-fly string method for minimum free energy paths calculation. *Chem. Phys. Lett.* **2007**, *446* (1−3), 182–190.

(20) Wienan, E.; Ren, W.; Vanden-Eijnden, E. Simplified and improved string method for computing the minimum energy paths in barrier-crossing events. *J. Chem. Phys.* **2007**, *126* (16), 164103.

(21) Khavrutskii, I. V.; Arora, K.; Brooks, C. L., III. Harmonic Fourier Beads Method for Studying Rare Events on Rugged Energy Surfaces. *J. Chem. Phys.* **2006**, *125* (17), 174108.

(22) Khavrutskii, I. V.; McCammon, J. A. Generalized gradient-augmented harmonic Fourier beads method with multiple atomic and/or center-of-mass positional restraints. *J. Chem. Phys.* **2007**, *127* (12), 124901.

(23) Khavrutskii, I. V.; Dzubiella, J.; McCammon, J. A. Computing Accurate Potentials of Mean Force in Electrolyte Solutions with the Generalized Gradient-Augmented Harmonic Fourier Beads Method. *J. Chem. Phys.* **2008**, *128* (4), 044106.

(24) Hu, H.; Lu, Z.; Parks, J. M.; Burger, S. K.; Yang, W. Quantum mechanics/molecular mechanics minimum free-energy path for accurate reaction energetics in solution and enzymes: Sequential sampling and optimization on the potential of mean force surface. *J. Chem. Phys.* **2008**, *128* (3), 034105.

(25) Burger, S. K.; Yang, W. Sequential quadratic programming method for determining the minimum energy path. *J. Chem. Phys.* **2007**, *127* (16), 164107.

(26) Burger, S. K.; Yang, W. Quadratic string method for determining the minimum-energy path based on multiobjective optimization. *J. Chem. Phys.* **2006**, *124* (5), 054109.

(27) Barducci, A.; Bussi, G.; Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* **2008**, *100* (2), 020603/1–020603/4.

(28) Branduardi, D.; Gervasio, F. L.; Parrinello, M. From A to B in free energy space. *J. Chem. Phys.* **2007**, *126* (5), 054103/1–054103/10.

(29) Laio, A.; Parrinello, M. Computing free energies and accelerating rare events with metadynamics. *Lecture Notes in Physics* **2006**, *703* (Computer Simulations in Condensed Matter Systems: From Materials to Chemical Biology, Volume 1), 315–347.

(30) Bussi, G.; Laio, A.; Parrinello, M. Equilibrium Free Energies from Nonequilibrium Metadynamics. *Phys. Rev. Lett.* **2006**, *96* (9), 090601/1–090601/4.

(31) Laio, A.; Rodriguez-Fortea, A.; Gervasio, F. L.; Ceccarelli, M.; Parrinello, M. Assessing the Accuracy of Metadynamics. *J. Phys. Chem. B* **2005**, *109* (14), 6714–6721.

(32) Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99* (20), 12562–12566.

(33) van der Vaart, A.; Karplus, M. Minimum free energy pathways and free energy profiles for conformational transitions based on atomistic molecular dynamics simulations. *J. Chem. Phys.* **2007**, *126* (16), 164106.

(34) Babin, V.; Roland, C.; Darden, T. A.; Sagui, C. The free energy landscape of small peptides as obtained from metadynamics with umbrella sampling corrections. *J. Chem. Phys.* **2006**, *125* (20), 204909.

(35) Babin, V.; Roland, C.; Sagui, C. Adaptively biased molecular dynamics for free energy calculations. *J. Chem. Phys.* **2008**, *128* (13), 134101.

(36) Huber, T.; Torda, A. E.; van Gunsteren, W. F. Local elevation: A method for improving the searching properties of molecular dynamics simulation. *J. Comput.-Aided Mol. Des.* **1994**, *8* (6), 695–708.

(37) Fukui, K. The Path of Chemical Reactions - The IRC Approach. *Acc. Chem. Res.* **1981**, *14* (12), 363–368.

(38) Fukui, K.; Kato, S.; Fujimoto, H. Constituent Analysis of the Potential Gradient Along a Reaction Coordinate. Method and an Application to CH4+T Reaction. *J. Am. Chem. Soc.* **1975**, *97* (1), 1–7.

(39) Hummer, G.; Szabo, A. Free Energy Surfaces from Single-Molecule Force Spectroscopy. *Acc. Chem. Res.* **2005**, *38* (7), 504–513.

(40) Kaestner, J.; Thiel, W. Bridging the gap between thermodynamic integration and umbrella sampling provides a novel analysis method: "Umbrella integration". *J. Chem. Phys.* **2005**, *123* (14), 144104.

(41) Cordero-Morales, J. F.; Cuello, L. G.; Zhao, Y.; Jogini, V.; Cortes, D. M.; Roux, B.; Perozo, E. Molecular determinants of gating at the potassium-channel selectivity filter. *Nat. Struct. Mol. Biol.* **2006**, *13* (4), 311–318.

(42) Domene, C.; Grottesi, A.; Sansom, M. S. Filter Flexibility and Distortion in a Bacterial Inward Rectifier K+ Channel: Simulation Studies of KirBac1.1. *Biophys. J.* **2004**, *87* (1), 256–267.

(43) Capener, C. E.; Proks, P.; Ashcroft, F. M.; Sansom, M. S. Filter Flexibility in a Mammalian K+ Channel: Models and SImulations of Kir6.2 Mutants. *Biophys. J.* **2003**, *84* (4), 2345–2356.

(44) Khalili-Araghi, F.; Tajkhorshid, E.; Schulten, K. Dyanmics of K+ Ion Conduction through Kv1.2. *Biophys. J.* **2006**, *91* (6), L72–L74.

(45) Hovmoeller, S.; Zhou, T.; Ohlson, T. Conformations of aminoacids in proteins. *Acta Crystallogr.* **2002**, *D58*, 768–776.

(46) MacKerell, A. D., Jr.; Feig, M.; Brooks, C. L., III. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **2004**, *25* (11), 1400–1415.

(47) Lee, M. S.; Feig, M.; Salsbury, F. R., Jr; Brooks, C. L., III. New analytic approximation to the standard molecular volume definition and its application to generalized Born calculations. *J. Comput. Chem.* **2003**, *24* (11), 1348–1356.

(48) Pettitt, B. M.; Karplus, M. The potential of mean force surface for the alanine dipeptide in aqueous solution: a theoretical approach. *Chem. Phys. Lett.* **1985**, *121* (3), 194–201.

(49) Lau, W. F.; Pettitt, B. M. Conformations of the Glycine Dipeptide. *Biopolymers* **1987**, *26* (11), 1817–1831.

(50) Long, S. B.; Tao, X.; Campbell, E. B.; MacKinnon, R. Atomic structure of a voltage-dependent K+ channel in a lipid membrane-like environment. *Nature* **2007**, *450* (7168), 376–383.

(51) Saparov, S. M.; Pohl, P. Beyond the diffusion limit: Water flow through the empty bacterial potassium channel. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101* (14), 4805–4809.

(52) Milner-White, E. J.; Watson, J. D.; Qi, G.; Hayward, S. Amyloid Formation May Involve alpha- to beta-Sheet Interconversion via Peptide Plane Flipping. *Structure* **2006**, *14* (9), 1369–1376.

(53) Armen, R. S.; DeMarco, M. L.; Alonso, D. O. V.; Daggett, V. Pauling and Corey's {alpha}-pleated sheet structure may define the prefibrillar amyloidogenic intermediate in amyloid disease. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101* (32), 11622–11627.

(54) Armen, R. S.; Alonso, D. O. V.; Daggett, V. Anatomy of an Amyloidogenic Intermediate: Conversion of beta-Sheet to alpha-Sheet Structure in Transthyretin at Acidic pH. *Structure* **2004**, *12* (10), 1847–1863.

(55) JiJi, R. D.; Balakrishnan, G.; Hu, Y.; Spiro, T. G. Intermediacy of Poly(L-proline) II and beta-Strand Conformations in Poly(L-lysine) beta-Sheet Formation by Temperature-Jump/UV Resonance Raman Spectroscopy. *Biochemistry* **2006**, *45* (1), 34–41.

(56) Asher, S. A.; Mikhonin, A. V.; Bykov, S. UV Raman Demonstrates that α-helical Polyalanine Peptides Melt to Polyproline II Conformations. *J. Am. Chem. Soc.* **2004**, *126* (27), 8433–8440.

(57) Heginbotham, L.; Lu, Z.; Abramson, T.; MacKinnon, R. Mutations in the K+ Channel Signature Sequence. *Biophys. J.* **1994**, *66*, 1061–1067.

(58) Splitt, H.; Meuser, D.; Borovok, I.; Betzler, M.; Schrempf, H. Pore mutations affecting tetrameric assembly and functioning of the potassium channel KcsA from Streptomyces lividans. *FEBS Lett.* **2000**, *472*, 83–87.

(59) MacKerell, A. D. Jr.; Bashford, D.; Bellott, M., Jr.; 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. All-atom empirical potential for molecular modeling and dynamics Studies of proteins. *J. Phys. Chem. B* **1998**, *102* (8), 3586–3616.

(60) MacKerell, A. D., Jr.; Brooks, B. R.; Brooks, C. L., III; Nilsson, L.; Roux, B.; Won, Y.; Karplus, M. , *CHARMM: The Energy Function and Its Parameterization with an Overview of the Program. In The Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Schreiner, P. R., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P., Henry, F., Schaefer, I., Eds.; John Wiley & Sons: Chichester, 1998.

(61) Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **1988**, *38* (6), 3098–3100.

(62) Becke, A. D., III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98* (7), 5648–5652.

(63) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37* (2), 785–789.

(64) Lee, M. S.; Salsbury, F. R., Jr.; Brooks, C. L., III. Novel generalized Born methods. *J. Chem. Phys.* **2004**, *116* (24), 10606–10614.

(65) Cossi, M.; Barone, V.; Mennucci, B.; Tomasi, J. Ab initio study of ionic solutions by a polarizable continuum dielectric model. *Chem. Phys. Lett.* **1998**, *286* (3−4), 253–260.

(66) Cances, E.; Mennucci, B.; Tomasi, J. A new integral equation formalism for the polarizable continuum model: theoretical background and applications to isotropic and anisotropic dielectrics. *J. Chem. Phys.* **1997**, *107* (8), 3032–3041.

(67) Mennucci, 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* (12), 5151–5158.

(68) Cossi, M.; Scalmani, G.; Rega, N.; Barone, V. New developments in the polarizable continuum model for quantum mechanical and classical calculations on molecules in solution. *J. Chem. Phys.* **2002**, *117* (1), 43–54.

(69) Ciccotti, G.; Ferrario, M.; Hynes, J. T.; Kapral, R. Constrained molecular dynamics and the mean potential for an ion pair in a polar solvent. *Chem. Phys.* **1989**, *129*, 241–251.

(70) 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* (8), 1011–21.

(71) Boczko, E. M.; Brooks, C. L., III. Constant-temperature free energy surfaces for physical and chemical processes. *J. Phys. Chem.* **1993**, *97* (17), 4509–13.

(72) Koslover, E. F.; Wales, D. J. Comparison of double-ended transition state search methods. *J. Chem. Phys.* **2007**, *127* (13), 134102.

(73) Weinan, E.; Ren, W.; Vanden-Eijnden, E. Finite Teperature String Method for the Study of Rare Events. *J. Phys. Chem. B* **2005**, *109* (14), 6688–6693.

(74) Maragliano, L.; Fischer, A.; Vanden-Eijnden, E.; Ciccotti, G. String method in collective variables: Minimum free energy paths and isocommittor surfaces. *J. Chem. Phys.* **2006**, *125* (2), 024106.

(75) Weinan, E.; Ren, W.; Vanden-Eijnden, E. String method for the study of rare events. *Phys. Rev. B* **2002**, *66*, 052301:1–4.

(76) Elber, R. Long-timescale simulation methods. *Curr. Opin. Struct. Biol.* **2005**, *15*, 151–156.

(77) Elber, R.; Cárdenas, A.; Ghosh, A.; Stern, H. Bridging the gap between long time trajectories and reaction pathways. *Adv. Chem. Phys.* **2003**, *126*, 93–129.

(78) Olender, R.; Elber, R. Calculation of classical trajectories with very large time step: Formalism and numerical examples. *J. Chem. Phys.* **1996**, *105* (20), 9299–9315.

(79) Khavrutskii, I. V.; Byrd, R. H.; Brooks, C. L., III. A line integral reaction path approximation for large systems via nonlinear constrained optimization: Application to alanine dipeptide and beta-haripin of protein G. *J. Chem. Phys.* **2006**, *124* (19), 194903.

(80) Quapp, W.; Heidrich, D. Analysis of the concept of minimum energy path on the potential energy surface of chemically reacting systems. *Theor. Chim. Acta* **1984**, *66* (3−4), 245–260.

(81) Sana, M.; Reckinger, G.; Leroy, G. An Internal Coordinate Invariant Reaction Pathway. *Theor. Chem. Acc.* **1981**, *58* (2), 145–153.

(82) Basilevsky, M. V. Modern Development of the Reaction Coordiante Concept. *J. Mol. Struct. (Theochem)* **1983**, *103* (12), 139–152.

(83) Lazaridis, T.; Tobias, D. J.; Brooks, C. L., III; Paulaitis, M. E. Reaction paths and free energy profiles for conformational transitions: an internal coordinate approach. *J. Chem. Phys.* **1991**, *95* (10), 7612–7625.

(84) Hirsch, M.; Quapp, W. Reaction pathways and convexity of the potential energy surface: applicaitons of Newton trajectories. *J. Math. Chem.* **2004**, *36* (4), 307–340.

(85) Maragliano, L.; Vanden-Eijnden, E. On-the-fly string method for minimum free energy paths calculation. *Chem. Phys. Lett.* **2007**, *446*, 182–190.

(86) Peters, B.; Heyden, A.; Bell, A. T.; Chakraborty, A. A growing string method for determining transition states: Comparison to the nudged elastic band and string methods. *J. Chem. Phys.* **2004**, *120* (17), 7877–7886.

(87) Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. *Numerical Recipes in Fortran 77: The Art of Scientific Computing;* Cambridge University Press: Port Chester, NY, 2001; Vol. 1.

(88) Cho, A. E.; Doll, J. D.; Freeman, D. L. The construction of double-ended classical trajectories. *Chem. Phys. Lett.* **1994**, *229* (3), 218–224.

(89) Eckart, C. Some studies concerning rotating axes and polyatomic molecules. *Phys. Rev.* **1935**, *47*, 552–558.

(90) Kudin, K. N.; Dymarsky, A. Y. Eckart axis conditions and the minimization of the root-mean-square deviation: Two

closely related problems. *J. Chem. Phys.* **2005**, *122* (22), 224105:1–2.

(91) Kabsch, W. A solution for the best rotation to relate two sets of vectors. *Acta Crystallogr.* **1976**, *A32*, 922–923.

(92) Trzesniak, D.; Kunz, A.-P. E.; van Gunsteren, W. F. A Comparison of Methods to Compute the Potential of Mean Force. *ChemPhysChem* **2007**, *8*, 162–169.

(93) Yu, H.-A.; Roux, B.; Karplus, M. Solvation thermodynamics: An approach from analytic temperature derivatives. *J. Chem. Phys.* **1990**, *92* (8), 5020–5033.

(94) Mark, P.; Nilsson, L. Structure and Dynamics of Liquid Water with Different Long-Range Interaction Truncation and Temperature Control Methods in Molecular Dynamics Simulations. *J. Comput. Chem.* **2002**, *23* (13), 1211–1219.

(95) Mark, P.; Nilsson, L. Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *J. Phys. Chem. A* **2001**, *105* (43), 9954–9960.

(96) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79* (2), 926–935.

(97) Lamoureux, G., Jr.; Roux, B. A simple polarizable model of water based on classical Drude oscillators. *J. Chem. Phys.* **2003**, *119* (10), 5185–5197.

(98) Ishimura, K.; Pulay, P.; Nagase, S. A new parallel algorithm of MP2 energy calculations. *J. Comput. Chem.* **2006**, *27* (4), 407–413.

(99) Johnson, E. R.; Becke, A. D. A unified density-functional treatment of dynamical, nondynamical, and dispersion correlations. II. Thermochemical and kinetic benchmarks. *J. Chem. Phys.* **2008**, *128* (12), 124105/1–124105/3.

(100) Becke, A. D.; Johnson, E. R. A unified density-functional treatment of dynamical, nondynamical, and dispersion correlations. *J. Chem. Phys.* **2007**, *127* (12), 124108/1–124108/8.

(101) Becke, A. D.; Johnson, E. R. A density-functional model of the dispersion interaction. *J. Chem. Phys.* **2005**, *123* (15), 154101/1–154101/9.

(102) 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.1*; Gaussian, Inc.: Pittsburgh, PA, 2003.

(103) Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M.; Tanabe, K. The magnitude of the CH/pi interaction between benzene and some model hydrocarbons. *J. Am. Chem. Soc.* **2000**, *122* (15), 3746–3753.

(104) Ulitsky, A.; Elber, R. A new technique to calculate steepest descent paths in flexible polyatomic systems. *J. Chem. Phys.* **1990**, *92* (2), 1510.

CT800086S

JCTC Journal of Chemical Theory and Computation

Is Alanine Dipeptide a Good Model for Representing the Torsional Preferences of Protein Backbones?

Michael Feig*

Department of Biochemistry and Molecular Biology and Department of Chemistry, Michigan State University, East Lansing, Michigan 48824

Received May 7, 2008

Abstract: The conformational preference for different ϕ/Ψ backbone torsion angles is a key determinant of peptide and protein secondary structure. Often, dipeptides are used as models for understanding protein backbone dynamics and to derive force field parameters. Here, the question is examined to what extent the conformational preferences in dipeptides reflect the backbone dynamics in polypeptides and proteins and to what extent an alanine dipeptide-based backbone torsion parametrization can lead to accurate reproduction of amino acid dependent ϕ/Ψ preferences in protein structures. Results from a comparison of the analysis of Protein Data Bank (PDB) structures with long simulations of selected proteins and amino acid dipeptides suggest that a common alanine dipeptide-based torsion potential does in fact lead to excellent agreement between protein simulations and PDB structures. At the same time, the ϕ/Ψ preferences in the dipeptides are significantly different, suggesting that dipeptides are not good model systems for studying protein backbone dynamics.

Introduction

Protein structure and dynamics are essential determinants of biological function. The structure of most proteins consists of well-defined three-dimensional folds that are built from α-helical and β-sheet secondary structure elements with connecting turns and loops. The ability to form different secondary structure elements is primarily a reflection of the conformational flexibility of the polypeptide backbone. There are essentially two backbone degrees of freedom for each amino acid residue: the torsion angles ϕ (C−N−C$_\alpha$−C) and Ψ (N−C$_\alpha$−C−N). For nonglycine and nonproline residues, the well-known Ramachandran plot[1] of Ψ versus ϕ identifies two major minima: α$_R$ ($\phi = -60$, $\Psi = -50$) in the α basin and PPII/C5 ($\phi = -60$ to -170, $\Psi = 120-170$) in the β basin (see Figure 1), which correspond to α-helical and extended β-strand/-sheet secondary structures when repeated over multiple amino acid residues. Secondary minima with higher relative free energies at α$_L$ ($\phi = 50$, $\Psi = 50$) and C7$_{ax}$ ($\phi = 50$, $\Psi = -130$) are relevant in the formation of turns and loops. The conformational preferences of proline are restricted to α$_R$, PPII ($\phi = -60$, $\Psi = 140$), and C7$_{eq}$ (ϕ

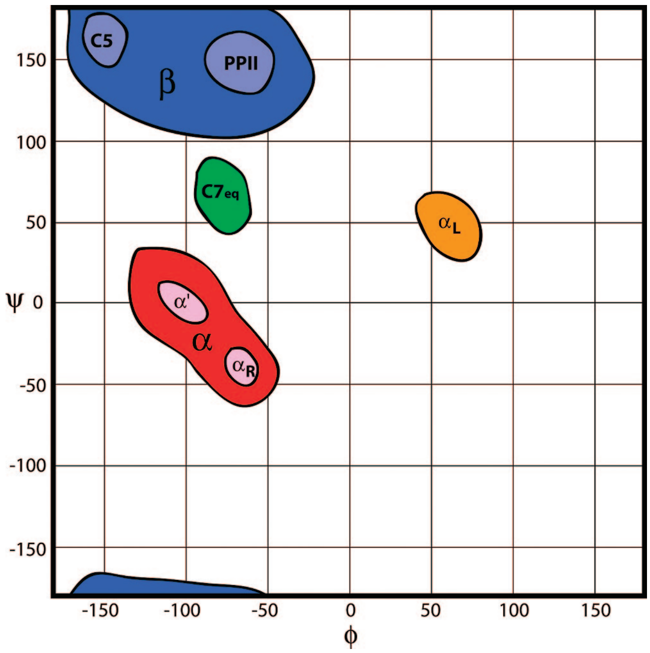


Figure 1. Schematic overview of major conformational basins sampled by ϕ/Ψ backbone torsion angles in nonglycine, nonproline peptide residues.

* Phone: (517) 432-7439. Fax: (517) 353-9334. E-mail: feig@msu.edu.

10.1021/ct800153n CCC: $40.75
Published on Web 07/30/2008

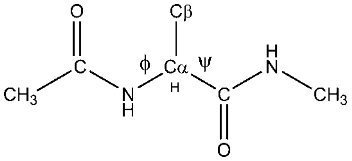


Figure 2. Blocked alanine dipeptide structure.

$= -75$, $\Psi = 75$) conformations. Glycine does not distinguish between positive and negative ϕ values due to its achiral nature and therefore visits the right side of the Ramachandran plot more frequently compared to the other amino acids. The preference for certain regions of the Ramachandran plot has resulted in the definition of "allowed" versus "disallowed" regions that are often applied in the validation of experimental and theoretical structures.[2,3] However, extensive analysis of high-resolution crystal structures has revealed that a significant fraction of amino acid residues may also exhibit ϕ/Ψ torsion angles outside those canonical regions.[3-5] Furthermore, the preferences for ϕ/Ψ torsions vary not just between proline and glycine but also among nonproline and nonglycine amino acid residues.[6,7] From detailed analyses in previous studies[6,7] and data gathered in this study, it has emerged that, in the β basin, alanine, tryptophan, phenylalanine, tyrosine, serine, glutamine, glutamic acid, arginine, lysine, methionine, and cysteine follow a similar energy landscape with two minima near C5 and PPII that are connected by a very shallow barrier. Valine, isoleucine, and to a lesser extent leucine have instead a single, broad minimum near ($\phi = -120$, $\Psi = 130$) that is intermediate between C5 and PPII and shifted downward to smaller Ψ angles. Aspartic acid, asparagine, and to a lesser degree histidine have an overall much broader β region that extends to $\Psi = 75$ and includes $C7_{eq}$. Finally, threonine exhibits four clearly discernible minima in the extended region, two at $\Psi = 165$ and two at $\Psi = 130$, all of which are connected by very shallow barriers. The energy landscape in the α_R basin also varies between different amino acids. Alanine, tryptophan, leucine, glutamine, arginine, glutamic acid, and methionine predominantly sample α_R ($\phi = -60$, $\Psi = -50$), while phenylalanine, tyrosine, asparagine, serine, threonine, histidine, lysine, aspartic acid, and cysteine also sample a second minimum at ($\phi = -100$, $\Psi = 0$) to a significant degree. Valine and isoleucine stand out by an extended low-energy region that includes ($\phi = -100$, $\Psi = 50$), which corresponds to π-helical conformations. These subtle but significant variations in ϕ/Ψ preferences can be rationalized in part by examining correlations with side-chain torsions.[7] Finally, a special case is given by amino acids that immediately precede proline.[3,8,9] These pre-Pro residues do not significantly populate conformations near ($\phi = -90$, $\Psi = 0$) in the α_R basin but instead populate so-called ζ conformations near ($\phi = -140$, $\Psi = 70$).

Blocked alanine dipeptide (see Figure 2) is commonly studied as a prototype of nonglycine/nonproline protein backbones since it allows full sampling of the ϕ/Ψ conformational space without the additional complexity of side-chain degrees of freedom. Numerous computational and experimental studies of alanine dipeptide have explored its thermodynamic,[10-14] kinetic,[13,15,16] and spectroscopic[17,18] properties. Furthermore, alanine dipeptide and sometimes alanine tri- or tetrapeptides[19] are commonly used for the testing and parametrization of amino acid backbones in molecular mechanics force fields.[19-21] Generally, a modular approach is followed, where alanine dipeptide-derived bonded and nonbonded parameters are used for modeling the backbone of all nonglycine and nonproline residues.[22] However, in some force fields, for example, Amber,[23] the partial charges of the backbone atoms may vary slightly for each amino acid.

The development of amino acid backbone parameters based on alanine dipeptide commonly relies on *ab initio* calculations since sufficiently detailed thermodynamic or kinetic data are not available from experiments. *Ab initio* calculations typically provide reference conformational energies in a vacuum for selected conformers. Recently, it has become possible to obtain conformational energies of alanine dipeptide from high-level theory over the entire range of ϕ/Ψ values on a grid with 15° resolution. This data has allowed much finer control in the parametrization of ϕ/Ψ torsion parameters and challenged the established paradigm of using a combination of univariate Fourier-series torsion potentials to generate the torsion potential.[21] In order to better represent the complex features of the ϕ/Ψ free energy landscape, a map-based spline-interpolated cross-correlation term (CMAP) has been introduced into the CHARMM force field.[21,24] The CMAP term can directly reproduce any given ϕ/Ψ map in alanine dipeptide and has been used in particular to reflect the vacuum conformational energies from the *ab initio* calculations. With the CMAP correction, the torsional preferences in peptides and proteins were found to be substantially improved by reducing an overemphasis on the sampling of π-helical structures[25] and by reducing deviations from crystallographic structures in molecular dynamics simulations.[24]

The possibility to exactly reproduce the *ab initio* ϕ/Ψ map of alanine dipeptide with the CMAP formalism raises the issue of whether parameters derived from alanine dipeptide in a vacuum are appropriate for all of the other (nonglycine and nonproline) amino acids and for condensed phase environments. More specifically, the question is whether the sampling of ϕ/Ψ torsion angles in protein simulations with a common underlying torsion potential reproduces the amino acid type-dependent variations found in crystallographic structures. A secondary point of biophysical interest is to what extent the ϕ/Ψ preferences observed for a given amino acid in the context of protein structures are inherently present at the dipeptide level or are a result of interactions due to the polypeptide and protein environments. In order to probe these questions, long-time molecular dynamics simulations of all amino acid dipeptides and selected proteins were carried out and compared with data extracted from crystallographic structures in the Protein Data Bank.[26] The results demonstrate that a single torsion potential is largely sufficient to reproduce the subtle variations in ϕ/Ψ preferences in the context of proteins. Furthermore, it is found that residue type-dependent variations in ϕ/Ψ preferences are largely absent at the dipeptide level and only fully materialize in the context of protein structures. The results are described and discussed

Table 1. Overview of Simulated Systems

system	starting structure	residues	box length [Å]	simulation length [ns]
dipeptides	extended	1	31.5−35.3	150.0
protein G	3GB1	56	61.0	50.0
ubiquitin	1D3Z	76	69.3	22.0
barnase	1A2P	108	56.9	148.0
barstar	1BTA	89	52.4	142.9
CheY	1CYE	129	56.4	124.7
FKBP12	1FKS	107	62.6	143.5
RNase A	2AAS	124	59.7	148.3
RNase H	2RN2	155	69.5	121.5

in more detail in the following, after a summary of the methodology used in this study.

Methods

Molecular dynamics simulations in explicit solvent were carried out for all amino acid dipeptides and eight small- to medium-size proteins. Dipeptides were blocked with an acetyl group at the N terminus and with N-methylamide at the C terminus. Each amino acid dipeptide was simulated in its standard protonation state at pH = 7. Two simulations were run for histidine, one protonated at N_δ, the other one at N_ε. In all cases except proline, the starting structure was a fully extended peptide with backbone torsions near ($\phi = -160$, $\Psi = 130$). The starting structure for proline was near ($\phi = -60$, $\Psi = 160$). The dipeptides were solvated with explicit water in a cubic box with at least 12 Å from any atom in the dipeptide to the closest edge of the box. Charged amino acids were neutralized with either a single chlorine or sodium ion, placed initially by randomly replacing one of the water molecules. Initial configurations were briefly minimized and then heated up to 298 K during a series of simulations with 1 ps at 50 K, 1 ps at 100 K, 1 ps at 150 K, 2 ps at 200 K, 2 ps at 250 K, 2 ps at 275 K, and 2 ps at 300 K. Further equilibration was carried out with three simulations at 300 K over 4 ps each. For each run, the average water density at the edge of the box was calculated and compared to the expected number density of water of 0.03337/Å^3 at 300 K and 1 atm of pressure. If any deviation was found, the box size was adjusted accordingly. Simulations in the NVT ensemble were then continued for another 150 ns to generate the production trajectories used for analysis. The CHARMM22 force field[20] with the CMAP torsion potential[21] and updated tryptophan parameters[27] was used for the dipeptide. Modified TIP3P[28] parameters from the CHARMM force field were used to model the explicit water. Ion parameters were taken from Roux.[29] Periodic boundaries were employed to avoid solvent boundary artifacts. Electrostatic interactions were calculated with the particle-mesh Ewald method[30] using a 32 × 32 × 32 grid for the discrete fast Fourier transform (FFT) and a 9 Å direct space cutoff. During the simulation, SHAKE[31] was applied to constrain the lengths of bonds involving hydrogen so that an integration time step of 2 fs could be used. The temperature was controlled with the Nosé−Hoover algorithm.[32]

Eight proteins were simulated over 22−148 ns. Table 1 summarizes the simulation details for each protein. Table 2 shows the number of each amino acid from the combined set of proteins. All proteins were started from experimental structures and solvated with sufficient counterions to neutralize each system. The same equilibration protocol and simulation parameters as described above for the dipeptide simulations were applied, except that larger FFT grid sizes were used according to the increased box sizes.

Table 2. Number of Each Amino Acid in Simulated Proteins and Protein Data Bank (PDB) Structures[a]

amino acid	Protein simulations	PDB chains
alanine	66	56716
arginine	35	34832
asparagine	37	28621
aspartic acid	49	39407
cysteine	14	8740
glutamine	38	25817
glutamic acid	57	46199
glycine	62	50196
histidine	10	15522
isoleucine	41	38399
leucine	63	62782
lysine	61	38784
methionine	14	14065
phenylalanine	19	27547
proline	27	32923
serine	43	39330
threonine	63	36648
tryptophan	15	10187
tyrosine	28	24099
valine	50	48320

[a] Pre-proline residues are not included in non-proline amino acid totals.

All of the simulations were run with the CHARMM program[33] in conjunction with the MMTSB Tool Set.[34] A trajectory analysis was also carried out with CHARMM and the MMTSB Tool Set.

The analysis of Protein Data Bank (PDB) structures was performed on the basis of 3326 chains from crystal structures with 2.0 Å resolution or better and not more than 25% sequence identity between any two chains. The list of chains was generated with the protein structure culling server PISCES[35] in June 2007. Table 2 shows the number of each amino acid in the analyzed PDB structures.

Results

Protein Simulations. The conformations sampled during the protein simulations were compared to experimental structures to gauge the degree of realism in the simulations. Experimental structures of monomeric, wild-type apo forms are available from both X-ray crystallography and NMR spectroscopy for all of the systems simulated here with one exception. The crystal structure of barstar was taken from the complex of barstar with ribonuclease Sa (PDB code: 1AY7). Average and final root-mean-square deviation (rmsd) values during the simulation as well as the rmsd of the average structure over the entire trajectory are reported in Table 3. The latter is the most appropriate measure when comparing to the experimental data. In general, the rmsd of the average is lower than the average instantaneous rmsd values. Furthermore, in all cases, the deviation from the crystallographic structures is less than the deviation from

Table 3. Root Mean Square Deviations from Experimental Structures in Protein Simulations (Standard Deviations Are Given in Parentheses)

system	reference	type	avg. C_α rmsd [Å]	C_α rmsd of final structure [Å]	C_α rmsd of avg. structure [Å]
protein G	3GB1	NMR	1.06(0.20)	1.43	0.79
	1PGB	X-ray	0.81(0.21)	0.84	**0.41**
ubiquitin	1D3Z	NMR	1.41(0.20)	1.28	1.25
	1UBQ	X-ray	1.24(0.18)	1.13	**1.04**
barnase	1FW7	NMR	1.71(0.15)	1.67	1.35
	1A2P	X-ray	1.54(0.25)	1.37	**1.15**
barstar	1BTA	NMR	1.34(0.16)	1.21	1.15
	1AY7B	X-ray	0.97(0.12)	0.85	**0.67**
CheY	1CYE	NMR	1.43(0.20)	1.70	1.18
	3CHY	X-ray	1.13(0.17)	1.14	**0.84**
FKBP12	1FKS	NMR	3.58(0.74)	4.77	2.74
	1FKK	X-ray	3.58(0.63)	4.58	**2.68**
RNase A	2AAS	NMR	2.49(0.43)	3.21	2.04
	8RAT	X-ray	2.18(0.34)	2.70	**1.58**
RNase H	1RCH	NMR	2.78(0.17)	2.89	2.54
	2RN2	X-ray	1.98(0.23)	2.01	**1.62**

the NMR structure. The rmsd of the average from the crystallographic structure is less than 1 Å for three of the proteins studied here and between 1 and 2 Å for four other systems. For FKBP12, the deviation is larger, 2.68 Å, due to large fluctuations of residues 32−45 and 80−95, which consist mostly of long loop regions. It is likely that even 150 ns is not sufficient to fully sample the conformational space of those flexible regions and that much longer simulations might be required to improve the agreement with the experimental structures that are averaged over much longer time scales and over a large number of molecules. While the small deviations of the average simulated structures from the experimental structures indicate a high level of realism in the simulations, the larger average instantaneous rmsd values with significant standard deviations indicate broad conformational sampling well beyond the time- and ensemble-averaged experimental structures.

ϕ/Ψ Sampling in Protein Simulations versus PDB Structures. The distribution of ϕ/Ψ backbone torsion angles was analyzed from the protein simulations as a function of the amino acid type and compared to the distributions from PDB structures. Results for selected amino acids representative of major variations in ϕ/Ψ sampling are shown in Figures 3 and 4. Data for all of the other amino acids are given in Figure S1 in the Supporting Information. The agreement between the results from the simulations and from the PDB is remarkably good, especially in the lower-energy regions. A prominent difference is the significant population of high-energy regions in the simulations from instantaneous conformational sampling over very long simulations. Most of these regions are populated only sparsely in the PDB structures.

From a more detailed comparison of the PDB distributions with the simulation results, it can be seen that the subtle variations as a function of amino acid type are reproduced well. In particular, there is good agreement in the following key features: In asparagine, the β region is more extended, the transition region between the α and β basins is lowered, the preference for a second minimum in the α-helical basin at ($\phi = -100$, $\Psi = 0$) is more pronounced, and the sampling of α_L conformations is relatively favorable. In threonine, the β region is split into four distinct minima and the preference for ($\phi = -100$, $\Psi = 0$) is enhanced and extended toward ($\phi = -140$, $\Psi = -30$). Finally, in valine, there is a broad minimum near ($\phi = -120$, $\Psi = 130$) and sampling of fully extended conformations near ($\phi = -180$, $\Psi = 180$) is reduced while the α-helical basin extends to ($\phi = -140$, $\Psi = -20$) and ($\phi = -100$, $\Psi = -50$).

It is remarkable how well the sequence-dependent variations in ϕ/Ψ preferences are reproduced with a single alanine−dipeptide-based CMAP torsion angle term, but there are also some deficiencies that could possibly be addressed through force field adjustments: In general, it appears that fully extended conformations near C5 are slightly too favorable over PPII conformations. Furthermore, sampling of the $C7_{eq}$ conformation near ($\phi = -75$, $\Psi = 75$) in the α/β transition region appears to be too unfavorable, which is especially apparent in alanine and asparagine. In asparagine and to a lesser extent in alanine, there is a third minimum in the simulations near ($\phi = -160$, $\Psi = 0$) which is not seen in the PDB distributions. Finally, valine did not sample the right side of the Ramachandran plot in the simulations. However, it is likely that this may be a reflection of the limited set of simulated structures rather than inherent force field deficiencies since a valine residue not initially found in a conformation with positive ϕ angles is unlikely to be able to assume such a conformation without major structural disruption unless it is located in a flexible loop region.

The backbone conformational preferences of proline and glycine residues are compared in Figure 4. It should be noted that different CMAP torsion potentials are used for those residues in the CHARMM force field to separately reproduce the *ab initio* ϕ/Ψ maps for proline and glycine dipeptide. The overall features of both maps are reproduced well between the simulations and PDB distributions, although there are also some notable differences: In glycine, there appears to be a lack of a clear minimum at α_R in the simulations. Instead, there is a minimum at ($\phi = -80$, $\Psi = 10$). There are also minima at ($\phi = -180$, $\Psi = -25$) and ($\phi = -160$, $\Psi = 30$) next to a high-energy region that do not seem to match the ϕ/Ψ preferences in the experimental structures, while the

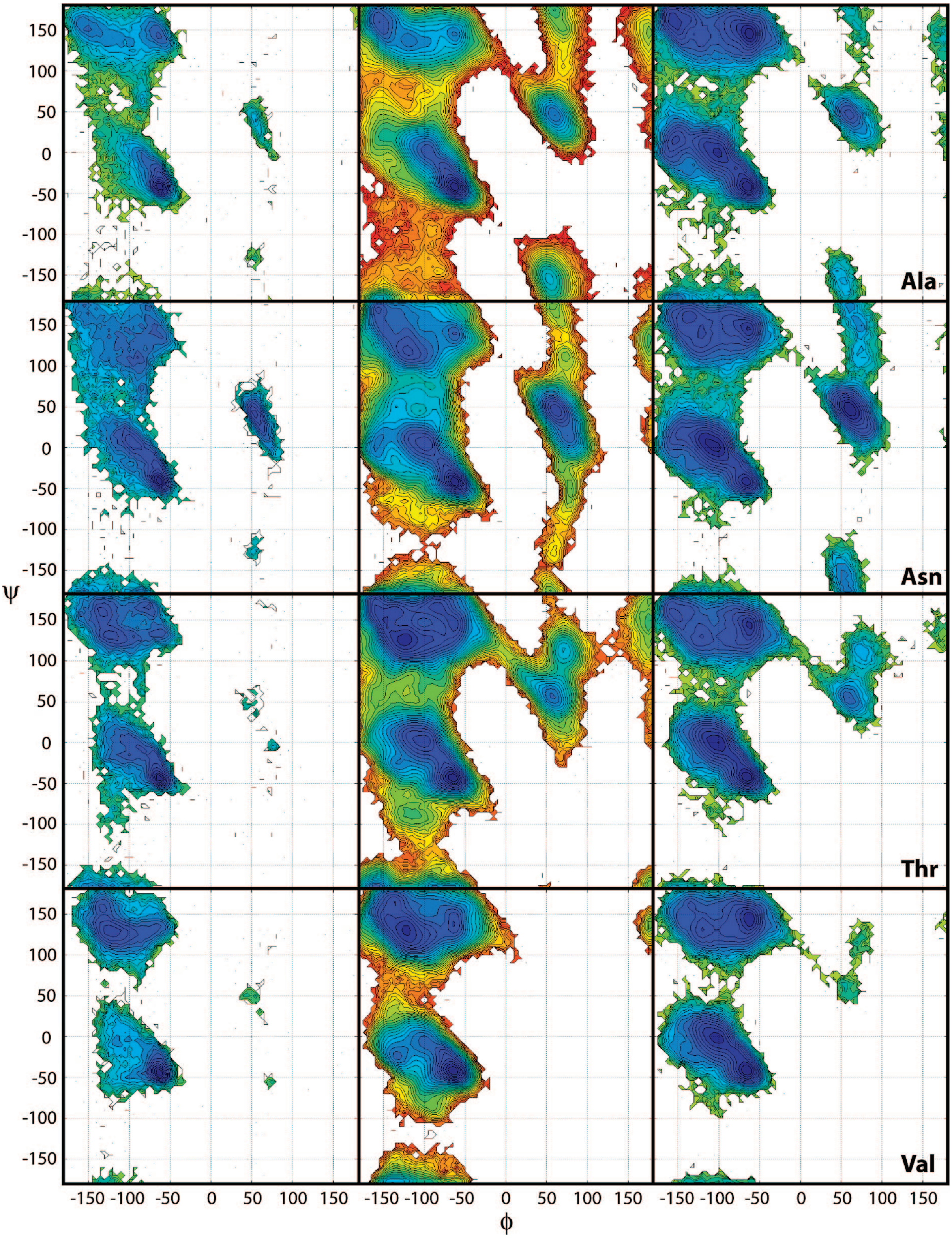


Figure 3. Potentials of mean force for the sampling of ϕ/Ψ backbone torsion angles in selected amino acid residues from PDB structures (left column), protein simulations (center column), and dipeptide simulations (right column). A color bar indicating the energy levels is given in Figure 4.

relative energy of C7$_{eq}$ conformations ($\phi = -75$, $\Psi = 75$) appears to be too high. In proline, the two major minima at α_R and PPII are reproduced reasonably well, but the C7$_{eq}$ conformation is again not favorable enough. Furthermore, the entire transition region is shifted to more negative ϕ angles.

Further analysis was carried out to compare the relative sampling of conformations in the major basins (α, β, and α_L). Table 4 shows the results for nonglycine and nonproline amino acids. The preference for sampling in the α basin versus the β basin matches to a large extent known secondary structure propensities,[36−38] especially for alanine and glutam-

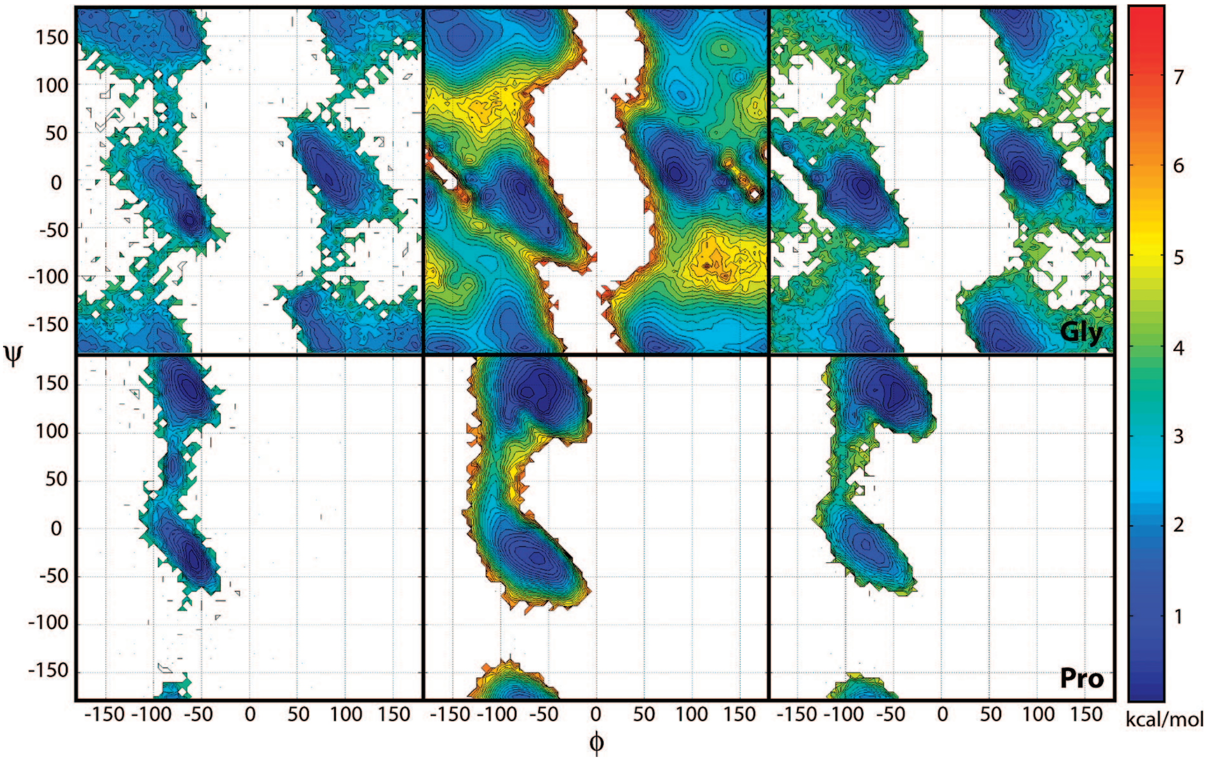


Figure 4. Potentials of mean force for the sampling of ϕ/Ψ backbone torsion angles in glycine and proline as in Figure 3.

Table 4. Relative Sampling (in %) of Different Regions in the Ramachandaran Plot for Each Amino Acid in PDB Structures, Simulated Proteins, and Dipeptides[a]

	PDB chains			protein simulations			dipeptides		
amino acid	α (α_R/α')	β	α_L	α (α_R/α')	β	α_L	α (α_R/α')	β	α_L
Ala	**66** (8.1)	31	1.2	**73** (4.4)	23	**3.6**	48 (0.80)	48	2.7
Arg	**60** (4.8)	36	2.5	**52** (3.3)	43	**4.4**	47 (0.68)	**50**	**3.0**
Asn	**51** (1.4)	33	**12**	**51** (1.7)	32	**16**	46 (0.28)	31	**21**
Asp	**57** (2.1)	34	**5**	**61** (2.2)	37	0.0	48 (1.41)	49	2.7
Cys	45 (3.2)	**51**	2.2	28 (3.4)	**67**	2.1	42 (0.59)	**51**	**6.4**
Gln	**64** (4.9)	31	2.7	**59** (4.3)	34	**6.1**	46 (0.57)	45	**8.4**
Glu	**68** (6.3)	29	1.7	**55** (4.1)	39	**4.7**	38 (1.46)	**60**	2.4
His	**52** (2.4)	41	**4.0**	23 (12)	29	**48**	36 (0.44)	42	**20**
Ile	47 (9.0)	**53**	0.1	33 (6.2)	**67**	0.2	49 (0.40)	**51**	0.2
Leu	**61** (5.5)	37	0.7	**57** (3.9)	42	0.0	46 (1.05)	**51**	2.3
Lys	**62** (5.1)	34	**3.0**	48 (6.4)	45	**5.8**	42 (0.79)	48	**8.8**
Met	**58** (5.2)	39	1.6	48 (6.5)	45	**5.2**	**54** (0.95)	45	1.1
Phe	49 (3.5)	48	1.7	23 (115)	**75**	0.3	42 (0.46)	46	**10**
Ser	**51** (2.6)	45	1.7	44 (1.6)	**50**	**4.9**	33 (0.69)	46	**19**
Thr	49 (2.2)	**50**	0.4	31 (1.3)	**67**	1.2	**57** (0.69)	41	1.9
Trp	**53** (4.5)	45	1.4	46 (12)	39	**14**	**51** (0.54)	46	2.5
Tyr	49 (3.1)	48	1.7	34 (1.2)	**61**	**3.9**	**52** (0.39)	38	**9.1**
Val	42 (7.9)	**58**	0.2	34 (7.8)	**66**	0.0	**55** (0.84)	45	0.1

[a] Pre-proline residues are not included in non-proline amino acid totals. The α basin is defined by the rectangle spanned by ($\phi = -180$, $\Psi = 50$) and ($\phi = 0$, $\Psi = -100$) with the α_R minimum at ($\phi = -80$, $\Psi = -15$) to ($\phi = -35$, $\Psi = -60$) and the secondary minimum (α′) at ($\phi = -140$, $\Psi = 40$) to ($\phi = -60$, $\Psi = -15$). The β basin is defined by ($\phi = -210$, $\Psi = 210$) to ($\phi = 0$, $\Psi = 85$) without further subdivision because of a wide variation in sampling in different amino acids. The α_L basin is defined by ($\phi = 0$, $\Psi = 100$) to ($\phi = 110$, $\Psi = -25$). α/β propensities larger than 50% and α_L propensities larger than 3% are highlighted in bold.

ic acid, which have high helical propensities.[36] However, it is interesting that the relative sampling of α_R versus α′ conformations within the α basin seems to be an overall better predictor of the propensity to form α helices. All of the amino acids with significant propensities to form α helices[36] (Ala, Gln, Glu, Ile, Leu, Lys, Met, Phe, Trp, and Val) strongly favor α_R sampling over α′ compared to the remaining residues, with arginine being the only exception. Conformations on the right-hand side of the Ramachandran plot are important in the formation of turns. The propensity to form α_L conformations is exceptionally high in asparagine and significant in aspartic acid, histidine, and lysine. This correleates with the high frequency of asparagine and aspartic acid residues in turn regions.[36]

The $\alpha/\beta/\alpha_L$ propensities from the protein simulations agree qualitatively with the results from the PDB analysis for most amino acids. Larger deviations are found for amino acids with a small number of representatives in the test sets, in particular cysteine, tryptophan, histidine, methionine, and phenylalanine, where the results presented here may not be statistically relevant. However, since only amino acids in flexible loop regions and at the termini are able to undergo conformational transitions between the major conformational basins without disrupting the overall structure, the simulations largely reflect the specific distribution of secondary structure elements in the simulated proteins rather than amino acid dependent propensities to form different secondary structure elements according to the underlying force field. The relative sampling of α_R versus α' conformations is also in good qualitative agreement between the simulations and PDB distributions (if amino acids that are rare in the simulations are excluded again). Overall, the ratio of α_R to α' sampling is smaller in the simulations (average, without Cys, Trp, His, Met, and Phe: $\alpha_R/\alpha' = 3.7$) compared to the PDB analysis (average $\alpha_R/\alpha' = 4.8$), suggesting that the sampling of α' might be too favorable in the simulations.

ϕ/Ψ Preferences in Dipeptides versus Proteins. The comparison of ϕ/Ψ preferences between the protein simulations and PDB structures provides an idea of how well the computational methodology can reproduce experimental data. On the other hand, a comparison of ϕ/Ψ preferences between protein simulations or experimental data and dipeptide simulations addresses the more fundamental question of to what extent amino acid dependent variations in ϕ/Ψ preferences found in proteins are already apparent at the dipeptide level. The results in Figure 3 show that the ϕ/Ψ preferences vary only to a small degree between different amino acid dipeptides, suggesting that amino acid dependent variations in ϕ/Ψ preferences do in fact stem predominantly from interactions due to polypeptide and protein environments. Closer inspection reveals some differences between different amino acids. Most significant are variations in the preferences for positive ϕ values. As in the protein simulations (and PDB distributions), asparagine dipeptide samples α_L more frequently, while valine dipeptide samples α_L less frequently than the alanine and threonine dipeptides. Furthermore, alanine and asparagine dipeptides (as well as arginine, cysteine, glutamine, glutamic acid, histidine, methionine, serine, tryptophan, and phenylalanine; see Figure S1, Supporting Information) extend the α basin toward ϕ values near -170, while the other dipeptides do not significantly populate that region. In the β basin, the overall minimum lies at PPII for all dipeptides, but very subtle variations in the conformational landscape of the β basin are apparent. These small differences, for example, diminished sampling near ($\phi = -170$, $\Psi = -170$) for valine, partially mimic the more pronounced variations in the β-basin landscape in the protein simulations and PDB structures but are far from completely reproducing the amino acid dependent variations seen in the protein context. Conformational preferences of glycine and proline dipeptides mostly resemble the preferences within the protein simulations, but differences in sampling fully extended conformations and the transition region near ($\phi = -100$, $\Psi = -100$) are apparent in glycine.

The relative sampling of the major conformational basins in the dipeptides also differs from the protein simulations and PDB structures (see Table 4). The relative sampling of the α basin is generally at or below 50% and less than the relative percentage in the PDB structures for most amino acids. Exceptions are threonine and valine, where conformations in the α basin are sampled more often in the dipeptide than in the protein context. The strong preference for α-helical conformations in alanine, glutamine, glutamic acid, leucine, and lysine found in the PDB structures is not apparent at the dipeptide level. It is particularly remarkable that glutamic acid, which is known to be a strong helix-forming amino acid,[36] actually has the highest propensity for extended structures at the dipeptide level compared to all of the other amino acids. Furthermore, the ratios of α_R to α' sampling are much lower in the dipeptides, mostly below 1, indicating that the sampling of α_R conformations is relatively disfavored in the dipeptides. Therefore, the polypeptide context and, in particular, the ability to form i, $i + 4$ backbone hydrogen bonding is essential in stabilizing α-helical secondary structure elements.

The α_L conformations are sampled at widely varying levels in the dipeptides. Asparagine, histidine, and serine dipeptides spend nearly 20% of the time in the α_L conformation, while isoleucine and valine essentially never sample α_L. This can be understood as a result of attractive intramolecular electrostatics between asparagine, histdine, and serine side chains and the peptide backbone and unfavorable side chain backbone interactions in the case of isoleucine and valine. The amino acid dependent propensities for α_L conformations in the dipeptides do not agree very well with the results from the protein simulations or PDB. However, an overall increased preference for α_L conformations compared to the PDB structures is apparent in both the dipeptide and protein simulations. This finding may suggest a need for raising the energy of α_L conformations in the force field.

Discussion and Conclusion

Previous studies have examined the detailed distribution of ϕ/Ψ torsion angles in experimental structures as a function of the amino acid type.[6,7] Here, these results are compared with torsional preferences from extensive simulations of proteins and dipeptides. The torsional preferences in the protein simulations are in good qualitative and quantitative agreement with the distribution of ϕ/Ψ angles found in PDB structures. Variations as a function of the amino acid type are generally represented well, including subtle features in the detailed energy landscape of the α and β basins. In contrast, ϕ/Ψ preferences in amino acid dipeptides vary much less as a function of the amino acid type. Some of the amino acid dependent variations seen in the context of proteins are also apparent at the dipeptide level, such as the preference for α_L sampling, but other features such as the preference for α-helical conformations in glutamic acid and the strong tendency to sample α_R conformations over α' are not reproduced in the dipeptide simulations. These results suggest that local interactions at the residue level play only a small role in determining the sequence-dependent torsional preferences of peptide backbones, while the more important contributions come from long-range interactions in the context of polypeptide chains and protein structures. An example is the observation of helix-capping interactions by glutamic acid residues that are not present at the dipeptide level.[39]

The same underlying CMAP torsion potential was used in all of the simulations. One of the main questions prompting this study is whether a common CMAP torsion potential for all nonglycine/nonproline residues is sufficient to accurately reproduce the sequence-dependent variations in ϕ/Ψ preferences. On the basis of the results presented here, this is apparently the case, further supporting the idea that the observed modulation of ϕ/Ψ preferences is largely a function of longer-range (electrostatic and Lennard-Jones type) interactions with neighboring residues and beyond.

Overall, the ϕ/Ψ preferences agree well between the protein simulations and PDB distributions. However, a close inspection suggests that the agreement could be improved further by slight force field adjustments. In particular, it appears that the sampling of positive ϕ values, of ϕ values below -150, and of the α' conformation is too favorable relative to other parts of the energy landscape, while $C7_{eq}$ sampling is underrepresented. There are also differences in the conformational preferences of glycine and proline residues that could be addressed by force field modifications. It is straightforward to adjust the CMAP torsion potentials accordingly, and future studies will examine how simulations with such a modified torsion potential affect the overall sampling of protein structures.

A constant concern with simulation studies is the achievement of converged sampling of all statistically relevant conformational regions. It appears that the dipeptide simulations over 150 ns are sufficient (or at least close to it) since many transitions are observed between the major basins in all of the simulations. However, it is possible that protein simulations of up to 150 ns do not completely sample the conformational space accessible during biological and experimental time scales of milliseconds to minutes. One consequence of the limited test sets is that the simulation results provide little information about the relative sampling of α versus β conformations since the observed α versus β propensities are largely a function of the native secondary structures of the chosen test proteins. Further studies of small helix- and hairpin-forming peptides with the same methodology will be necessary to examine the relative sampling of α versus β conformations in the context of proteins in more detail. However, the sampling of relative conformations within a given basin is expected to be more meaningful since the corresponding structural variations could largely be accommodated without major disruption of a given protein structure.

The dipeptide simulations can be compared to spectroscopic data that indicate that PPII and C5 are the dominant conformations in solution, while α_R is populated only to a small extent in alanine and valine dipeptides.[14,18] The dipeptide simulations presented here show a more equal sampling of α and β basins, suggesting a slight bias toward α-helical conformations. Such a bias has also been suspected in other recent studies with the CHARMM force field in conjunction with the CMAP potential and will require further exploration.[40] A force field that better reproduces experimental data for dipeptides and other small peptides may also alter the torsional preferences in the protein simulations reported here. The hope is that such modifications would improve the agreement with the conformational preferences from the PDB and lead to conformational sampling in even better agreement with crystallographic structures for individual proteins. It is possible, though, that the fixed charged force field used here places limitations on how well experimental data for small peptides and larger proteins can be reproduced simultaneously. In this context, it should be stressed that the simulations reported here only consider the combination of the CHARMM force field with the CMAP torsion potential, and specific results may vary for other force fields. It is likely, however, that the general conclusions are equally valid for other force fields where similar assumptions of a common backbone torsion potential based on alanine dipeptide are made.

Finally, it should be noted that there is a fundamental difference in the way the potentials of mean force are obtained from the simulations and from the experimental structures. The results from the simulations are obtained from a small number of structures over a large number of instantaneous conformations. On the other hand, the results from the experiment are obtained from a large number of structures, each representing an ensemble and time average. The potentials of mean force agree quantitatively very well in the low-energy regions, thereby confirming the validity of the ergodic hypothesis that time averages are equivalent to ensemble averages. In contrast, higher-energy regions are not sampled extensively in the PDB structures, while the simulations show broad conformational sampling well beyond the major conformational basins. This difference is primarily a result of the relatively small sample size used in the analysis of the PDB structures. For example, approximately 57 000 alanine conformations were analyzed from PDB structures (see Table 2) compared to approximately 17 million conformations from the simulations (66 alanine residues over an average simulation length of 130 ns with conformations saved every 0.5 ps). However, it is also possible that experimental structures, except for structures at the very highest resolution, reflect to some extent assumptions about ideal molecular bonding geometries if imposed during molecular refinement. Such constraints would limit the sampling of noncanonical regions of the Ramachandran plot in the experimental structures. It should be mentioned that there are also some theoretical concerns that have been raised about extracting potentials of mean force from PDB structures;[41] however, these arguments may not apply to the present study since we are analyzing simulations and crystallographic structures in an equivalent manner.

We now come back to the central question of this paper: Is alanine dipeptide a good model for representing the torsional preferences of protein backbones? The answer is "yes" and "no". It appears that force field parametrization based on alanine dipeptide along with suitable long-range interactions can accurately reflect amino acid dependent variations in backbone torsional preferences in the context of protein structures. This suggests that a modular approach in the development of the force field is justified, and specific modifications to bonding terms as a function of amino acid, with the exception of glycine and proline, are probably not necessary. However, the ϕ/Ψ preferences differ significantly between dipeptide and protein environments. As a consequence, dipeptides do not appear to be a suitable model for understanding the backbone torsional preferences of amino acids in proteins.

Acknowledgment. Financial support from NSF CAREER grant 0447799 and the Alfred P. Sloan Foundation and access to computational resources through the High Performance Computing Center at Michigan State University are acknowledged.

Supporting Information Available: Potentials of mean force for sampling of ϕ/ψ backbone torsion angles for remaining amino acid residues. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Ramachandran, G. N.; Ramakrishnan, C.; Sasisekharan, V. Stereochemistry of Polypeptide Chain Configurations. *J. Mol. Biol.* **1963**, *7*, 95–99.

(2) Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. PROCHECK: A program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **1993**, *26*, 283–291.

(3) Lovell, S. C.; Davis, I. W.; Arendall, W. B., III; de Bakker, P. I. W.; Word, J. M.; Prisant, M. G.; Richardson, J. S.; Richardson, D. C. Structure Validation by Cα Geometry: ϕ, Ψ and Cβ Deviation. *Proteins* **2003**, *50*, 437–450.

(4) Karplus, P. A. Experimentally observed conformation-dependent geometry and hidden strain in proteins. *Protein Sci.* **1996**, *5*, 1406–1420.

(5) Kleywegt, G. J.; Jones, T. A. Phi/Psi-chology: Ramachandran revisited. *Structure* **1996**, *4*, 1395–1400.

(6) Novmöller, S.; Zhou, T.; Ohlson, T. Conformations of amino acids in proteins. *Acta Crystallogr., Sect. D* **2002**, *58*, 768–776.

(7) Chakrabarti, P.; Pal, D. The interrelationships of side-chain and main-chain conformations in proteins. *Prog. Biophys. Mol. Biol.* **2001**, *76*, 1–102.

(8) Ho, B. K.; Brasseur, R. The Ramachandran plots of glycine and pre-proline. *BMC Struct. Biol.* **2005**, *5*, 14–24.

(9) Anderson, R. J.; Weng, Z.; Campbell, R. K.; Jiang, X. Main-Chain Conformational Tendencies of Amino Acids. *Proteins* **2005**, *60*, 679–689.

(10) Smith, P. E. The alanine dipeptide free energy surface in solution. *J. Chem. Phys.* **1999**, *111*, 5568–5579.

(11) Drozdov, A. N.; Grossfield, A.; Pappu, R. V. Role of Solvent in Determining Conformational Preferences of Alanine Dipeptide in Water. *J. Am. Chem. Soc.* **2004**, *126*, 2574–2581.

(12) Wang, Z. X.; Duan, Y. Solvation effects on alanine dipeptide: A MP2/cc-pVTZ//MP2/6-31G** study of (Phi,Psi) energy maps and conformers in the gas phase, ether, and water. *J. Comput. Chem.* **2004**, *25*, 1699–1716.

(13) Feig, M. Kinetics from Implicit Solvent Simulations of Biomolecules as a Function of Viscosity. *J. Chem. Theory Comput.* **2007**, *3*, 1734–1748.

(14) Kwac, K.; Lee, K. K.; Han, J. B.; Oh, K. I.; Cho, M. Classical and quantum mechanical/molecular mechanical molecular dynamics simulations of alanine dipeptide in water: Comparisons with IR and vibrational circular dichroism spectra. *J. Chem. Phys.* **2008**, *128*, 105106.

(15) Chekmarev, D. S.; Ishida, T.; Levy, R. M. Long-time conformational transitions of alanine dipeptide in aqueous solution: Continuous and discrete-state kinetic models. *J. Phys. Chem. B* **2004**, *108* (50), 19487–19495.

(16) Swope, W. C.; Pitera, J. W.; Suits, F.; Pitman, M.; Eleftheriou, M.; Fitch, B. G.; Germain, R. S.; Rayshubski, A.; Ward, T. J. C.; Zhestkov, Y.; Zhou, R. Describing protein folding kinetics by molecular dynamics simulations. 2. Example applications to alanine dipeptide and beta-hairpin peptide. *J. Phys. Chem. B* **2004**, *108* (21), 6582–6594.

(17) Kim, Y. S.; Wang, J. P.; Hochstrasser, R. M. Two-dimensional infrared spectroscopy of the alanine dipeptide in aqueous solution. *J. Phys. Chem. B* **2005**, *109* (15), 7511–7521.

(18) Grdadolnik, J.; Grdadolnik, S. G.; Avbelj, F. Determination of conformational preferences of dipeptides using vibrational spectroscopy. *J. Phys. Chem. B* **2008**, *112*, 2712–2718.

(19) Beachy, M. D.; Chasman, D.; Murphy, R. B.; Halgren, T. A.; Friesner, R. A. Accurate ab Initio Quantum Chemical Determination of the Relative Energetics of Peptide Conformations and Assessment of Empirical Force Fields. *J. Am. Chem. Soc.* **1997**, *119*, 5908–5920.

(20) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, J. D.; Evanseck, M. J.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(21) MacKerell, A. D., Jr.; Feig, M.; Brooks, C. L., III. Improved treatment of the protein backbone in empirical force fields. *J. Am. Chem. Soc.* **2004**, *126*, 698–699.

(22) Mackerell, A. D. Empirical force fields for biological macromolecules: Overview and issues. *J. Comput. Chem.* **2004**, *25* (13), 1584–1604.

(23) Case, D. A.; Cheatham, T. E., III; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber Biomolecular Simulation Programs. *J. Comput. Chem.* **2005**, *26* (16), 1668–1688.

(24) MacKerell, A. D., Jr.; Feig, M.; Brooks, C. L., III. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **2004**, *25*, 1400–1415.

(25) Feig, M.; MacKerell, A. D.; Brooks, C. L. Force field influence on the observation of pi-helical protein structures in molecular dynamics simulations. *J. Phys. Chem. B* **2003**, *107* (12), 2831–2836.

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

(27) Macias, A. T.; MacKerell, A. D. CH/pi interactions involving aromatic amino acids: Refinement of the CHARMM tryptophan force field. *J. Comput. Chem.* **2005**, *26* (14), 1452–1463.

(28) Jorgensen, W. L. Quantum and statistical mechanical studies of liquids. 10. Transferable intermolecular potential functions for water, alcohols, and ethers. Application to liquid water. *J. Am. Chem. Soc.* **1981**, *103*, 335–340.

(29) Roux, B. Valence selectivity of the gramicidin channel: A molecular dynamics free energy perturbation study. *Biophys. J.* **1996**, *71*, 3177–3185.

(30) Darden, T. A.; York, D.; Pedersen, L. G. Particle Mesh Ewald: An Nlog(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(31) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes. *J. Comput. Phys.* **1977**, *23* (3), 327–341.

(32) Nose, S. A Molecular Dynamics Method for Simulations in the Canonical Ensemble. *Mol. Phys.* **1984**, *52*, 255–268.

(33) 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.

(34) Feig, M.; Karanicolas, J.; Brooks, C. L., III. MMTSB Tool Set: Enhanced Sampling and Multiscale Modeling Methods for Applications in Structural Biology. *J. Mol. Graphics Modell.* **2004**, *22*, 377–395.

(35) Wang, G.; Dunbrack, R. L. PISCES: a protein sequence culling server. *Bioinformatics* **2003**, *19*, 1589–1591.

(36) Chou, P. Y.; Fasman, G. D. Prediction of protein conformation. *Biochemistry* **1974**, *13*, 222–245.

(37) Guzzo, A. The influence of amino-acid sequence on protein structure. *Biophys. J.* **1965**, *5*, 809–822.

(38) Lewis, P. N.; Go, N.; Go, M.; Kotelchuck, D.; Scheraga, H. A. Helix Probability Profiles of Denatured Proteins and Their Correlation with Native Structures. *Proc. Natl. Acad. Sci. U.S.A.* **1970**, *65*, 810–815.

(39) Stellwagen, E.; Shalongo, W. Evidence for Glutamate Self-Capping Within a Peptide Helix. *Biopolymers (Peptide Sci.)* **1997**, *43*, 413–418.

(40) Tanizaki, S.; Clifford, J. W.; Connelly, B. D.; Feig, M. Conformational Sampling of Peptides in Cellular Environments. *Biophys. J.* **2008**, *94*, 747–759.

(41) Ben-Naim, A. Statistical potentials extracted from protein structures: Are these meaningful potentials. *J. Chem. Phys.* **1997**, *107* (9), 3698–3706.

CT800153N