JCTC Journal of Chemical Theory and Computation

The Origin of Layer Structure Artifacts in Simulations of Liquid Water

David van der Spoel* and Paul J. van Maaren

Department of Cell and Molecular Biology, Uppsala University, Husargatan 3, Box 596, SE-751 24 Uppsala, Sweden

Received September 9, 2005

Abstract: A recent paper (Yonetani, *Chem. Phys. Lett.* **2005**, *406*, 49−53) shows that in computer simulations of TIP3P water (Jorgensen et al. *J. Chem. Phys.* **1983**, *79*, 926−935) a strange layer formation can occur when a long cutoff is used. This result is counterintuitive because, in principle, increasing the cutoff should give more accurate results. Here we test this finding for different water models and try to explain why layer formation occurs. In doing so we find that under certain conditions, layer formation coincides with a sharp density increase to 1050 g/L, while simultaneously a pressure of 600 bar develops and water diffusion becomes anisotropic. This leads us to conclude that a group-based cutoff (of at least 1.4 nm) stabilizes an anomalous phase with most water models. In some cases the ordering is strengthened further by periodicity in the simulation cell, but periodicity effects can even be observed with a short cutoff (0.9 nm) and a relatively large box of 4 nm. Water models that have a relatively large quadrupole moment, more in accord with the experimental gas-phase values, in particular TIP4P (Jorgensen et al. *J. Chem. Phys.* **1983**, *79*, 926−935), are much less affected by the problem, because the dipole−dipole interaction is quenched at long distance. A comparison of different cutoff treatments, namely truncation, reaction field, particle mesh Ewald (PME), and switch and shift functions, for the simulation of water shows that only PME and shift functions yield realistic dipole−dipole interactions at long distance. The impact for biomolecular simulations is discussed.

1. Introduction

In a recent paper, Yonetani[1] reported a strange artifact: when TIP3P water[2] is simulated with group-based truncation and a cutoff of 1.8 nm it forms liquid layers in which the molecular dipoles are mainly oriented in the same direction.[1] The artifact was reproduced in two different MD packages, Amber[3] and GROMACS,[4−6] implying that it is not due to a software bug. The layers were also found in larger boxes,[1] indicating that the behavior is a truncation problem rather than being related to the use of periodic boundary conditions. Interestingly, the layer formation does not occur in conjunction with the particle-mesh Ewald (PME)[7,8] method for computing electrostatic interactions. PME is one of the most popular implementations of the Ewald technique, because it is efficient and fast. The electrostatic energy is split into two parts, the short range is computed in direct space and the long range in reciprocal space, using fast Fourier transforms. Other algorithms in practical use for the computation of long-range Coulomb interactions are the particle−particle particle-mesh method,[9,10] fast multipole methods,[11,12] and Lekner summation.[13−16]

Here we present a series of simulations, using different water models and simulation conditions, in an attempt to explain the phenomenon. For a particular set of parameters we find that TIP3P water transforms to a state with a density of 1050 g/L, a pressure of 600 bar, and a potential energy lowered by 1 kJ/mol compared to the normal −40.8 kJ/mol.[17] Similar results are obtained for TIP5P, SPC/E, and SPC water. The simultaneous increase in density and pressure is driven by an increased number of hydrogen bonds and a clearly different, layered, structure; in addition diffusion

* Corresponding author phone: 46-18-4714205; fax: 46-18-511755; e-mail: spoel@xray.bmc.uu.se.

10.1021/ct0502256 CCC: $33.50
Published on Web 11/25/2005

becomes anisotropic. The combination of these observations leads to the conclusion that this is an artificial phase of liquid water.

Complex phenomena such as phase transitions are particularly sensitive to the correctness of the simulation conditions. Slovak and Tanaka have shown[18] in simulation studies of melting of ice VII that with a short (0.8655 nm) "smoothly" truncated potential the melting temperatures are very different from those obtained with Ewald summation. In other studies of phase transitions Zangi and Mark[19] used a twin-range cutoff with reaction field for the Coulomb interaction, while Matsumoto et al.[20] used a shifted potential as first described by Ohmine et al.[21] (Appendix B). From a personal communication we learned that Yamada et al.[22] used a plain cutoff of 0.9 nm (group-based truncation[23,24]), whereas Koga et al.[25] used a cutoff of 0.875 nm with a switching function. Since there is such a plethora of methods for cutoff treatment, it is very important that authors document their work sufficiently[26,27] to allow others to verify it. This work focuses on the treatment of electrostatic interactions, but in principle the accuracy of other simulation algorithms, like integrators and constraint treatment, need to be considered as well. The impact of those algorithms on accuracy fall outside the scope of this paper, however.

2. Methods

Molecular dynamics simulations were performed using the TIP3P and TIP4P water models,[2] the TIP5P model,[28] the SPC model,[29] and the SPC/E model.[30] Berendsen temperature coupling (298.15 K) and pressure coupling (1 bar) were used.[31] The temperature coupling constant τ_T was 0.1 ps, and the pressure coupling constant τ_P was 0.2 ps unless otherwise stated. A compressibility of 0.00005 (1/bar) was used. Although we realize that τ_P is unusually short, we used this value to be compatible with Yonetani;[1] in addition, we explicitly tested the influence of τ_P as described in the results. In all cases a cutoff was used for the Lennard-Jones interactions, but long-range corrections to the energy were applied in the standard way.[23] Neighborlists were used and updated every fifth integration time step, which was 2 fs. The water molecules were kept rigid using the SETTLE algorithm.[32] Center of mass motion of the simulation box was removed at every time step.[33]

Five different cutoff schemes were used: 1. a group-based cutoff (simple truncation at the indicated cutoff distance), 2. a reaction field[34−36] with $\epsilon_{rf} = 78.5$ (Appendix A), 3. the particle-mesh Ewald algorithm,[7,8] 4. an atom-based switch function, and 5. an atom-based shift function (Appendix B). For all simulations molecule-based neighbor searching was done, and for both the cutoff and reaction field it should be noted that the cutoff was based on molecules as well, to avoid artifacts due to non-neutral groups. The atom-based switch and shift functions go to zero smoothly; however, it is known that atom based switch functions can cause artifacts when the switching range is too short.[24] When using PME the grid-spacing was 0.12 nm (fluctuating slightly due to pressure coupling), and fourth-order B-splines were used for charge spreading and force interpolation. Conducting boundary conditions were used for PME.

Table 1. Properties of the Water Molecules Used: Dipole μ, the Components of the Quadrupole Tensor Θ, the Root Mean Square Deviation of the Quadrupole Tensor Elements from the Experimental Values

model	μ (D)	Θ_{xx} (10^{-1}D nm)	Θ_{yy} (10^{-1}D nm)	Θ_{zz} (10^{-1}D nm)	RMSD(Θ) (10^{-1}D nm)
expt	1.855	−2.50	2.63	−0.13	
TIP3P	2.35	−1.68	1.76	−0.08	1.20
TIP4P	2.18	−2.09	2.20	−0.11	0.59
TIP5P	2.29	−1.48	1.65	−0.17	1.42
SPC	2.274	−1.82	2.11	−0.29	0.87
SPC/E	2.351	−1.88	2.19	−0.30	0.78

An overview of the simulations performed and the conditions is given in Table 2. All simulations were 2 ns long unless otherwise stated. In total well over 100 simulations were performed, all using the GROMACS software.[4−6] All simulations used single-precision arithmetic, except one, which was performed in order to check the effect of precision.

3. Results

3.1. Dipole−Dipole Correlation. Analysis of the simulations focuses on dipole orientation; in particular, we look at the distance dependent Kirkwood factor $G_k(r)$[36] according to

$$G_k(r) = \sum_{r_{ij}<r} \frac{\mu_i \cdot \mu_j}{\mu^2} \tag{1}$$

where μ_i and μ_j are the dipole vectors of water molecules i and j, respectively, r_{ij} is the distance between oxygen atoms, and the dielectric constant $\epsilon(0)$ is determined from the fluctuations of the total system dipole.[37]

In Figure 1a we have plotted the cutoff dependence of $G_k(r)$ for TIP3P. Obviously there is a severe artifact around the cutoff distance. This is a well-known problem that has been described in detail before.[17,38] In Figure 1b the average cosine $\langle\cos\rangle$ of the angle between two molecular dipoles at a distance r is given. Here we can clearly see that there is a dip in the function corresponding to an anticorrelation on average around the cutoff distance. Obviously, the orientational correlation between water molecules should approach zero with increasing distance,[39,40] so this is an artifact due to the cutoff. In a further series of simulation of 10 648 TIP3P molecules we varied the cutoff r_c from 0.9 to 3.35 nm. The minimum of the $G_k(r)$ function becomes deeper with increasing r_c, indicating that ever stronger ordering is induced (Figure 1c). This shows that there is nothing special about the cutoff of 1.8 nm that was used by Yonetani.[1] In Figure 1d,e we have plotted the first minimum of the oxygen−oxygen and oxygen−hydrogen radial distribution function (RDF), respectively, for different cutoffs. These parts of the RDF were plotted as it is here that the main difference is visible. The minimum in oxygen−oxygen RDF is slightly less deep with long cutoffs, while it is the reverse for the oxygen hydrogen RDF. Apparently the water structure changes slightly upon increasing the cutoff, to accommodate the higher density (Figure 2).

Table 2. Overview of the Simulations Performed[a]

model(s)	N_{mol}	r_c (nm)	elect.	comment
TIP3P	820	0.9−1.4	cut	0.1 nm increments
TIP3P	2201	0.9−1.8	cut	0.1 nm increments
TIP3P	2201	1.8	cut	double precision
TIP3P	10648	0.9−3.3	twin	rc = 0.9 nm, in 0.1 nm increments
TIP3P	10648	3.05−3.35	twin	rc = 0.9 nm, in 0.1 nm increments
TIP3P SPC SPC/E	1728	0.9−1.8	cut	0.1 nm increments (1 ns)
TIP3P	17608	1.8	cut	
TIP3P	59427	1.8	cut	180 ps only
TIP3P	2201	1.8	cut	$\tau_P = 0$ (i.e. NVT), 0.5, 1.0, and 2.0 ps
TIP3P		1.8	cut	N_{mol} = 1728, 2197, 2744, 3375, 4096, 4913, 5832, 6859, 8000, 9261, 10648, 12167, 13824, 15625, 17576, 19683, 21952
TIP3P	2201	1.8	cut	Anisotropic pressure scaling with $\tau_P = 5$ ps. The box edges were allowed to vary independently.
TIP3P	2201	1.8	cut	Anisotropic pressure scaling with $\tau_P = 5$ ps. Only one of the box edges was allowed to vary, while the others were kept at their initial value of 4.06 nm.
TIP3P	2201	1.8	cut	dodecahedron box
All	2201	1.8	cut	
All	2201	1.8	RF	reaction field[34−36] with $\epsilon_{rf} = 78.5$
SPC SPC/E	2201	1.8	RF	reaction field with $\epsilon_{rf} = \infty$ and with the so-called "self-consistent" $\epsilon_{rf} = 54$ (SPC) and 62.3 (SPC/E) as determined by Smith and Van Gunsteren[38] as well as additional simulations with $\epsilon_{rf} = 65$ (SPC) and 77 (SPC/E).
All	2201	0.9	PME	particle-mesh Ewald method[7, 8]
All	2201	1.7	switch	switch function (see Appendix) with a cutoff of 1.7 nm, a 0.2 nm switching range (i.e. $r_1 = 1.5$, $r_c = 1.7$ nm), neighborlists were computed with a 1.8 nm cutoff
All	2201	1.7	switch	switch function (see Appendix) with a cutoff of 1.7 nm, a 0.4 nm switching range (i.e. $r_1 = 1.3$, $r_c = 1.7$ nm), neighborlists were computed with a 1.8 nm cutoff
All	2201	1.7	shift	shift function with a cutoff r_c of 1.7 nm, neighborlists were computed with a 1.8 nm cutoff
TIP3P	17608	0.9	PME	
TIP3P	2201	0.9	PME	dodecahedron box

[a] Water model, number of molecules, cutoff distance, and electrostatics treatment. "All" in column model indicates, TIP3P, TIP4P, TIP5P, SPC, and SPC/E.

The energy (Figure 2a) and density (Figure 2b) show a weird cutoff dependence, which is related to layer formation. For instance for $r_c = 1.5$ and 1.7 nm we obtain relatively large error bars (probably) due to intermittent layer formation. Figure 2c shows the dependence of the diffusion constant on the cutoff. Obviously, cutoffs of 1.4 nm and larger induce significant changes in water properties. In Figure 2d we plot the ratio of diffusion coefficients in a plane normal to the box axes and parallel to the box axis. For isotropic diffusion the ratio should be 1. Since the numbers are averaged over the whole simulation, further compensation effects are expected, but for $r_c = 1.7$ nm we see a very strong peak, which hence indicates that relatively stable layers should be present in the simulation. For the anisotropic simulation (described in more detail below) where we have stable layers perpendicular to the *X*-axis, we see that the ratio of diffusion coefficients perpendicular to and along the *X*-axis is less than one, meaning that the diffusion perpendicular to the layer structure is considerably faster than within the layer.

In the extensive series of simulations that we have performed, we found that in some cases the artifacts get even worse. In a particular case, when anisotropic pressure scaling was used in which only a single axis of the simulation box was allowed to fluctuate, very stable layer formation occurred. Figure 3 shows how several properties evolve in a simulation of 2201 TIP3P molecules under these conditions. Figure 3a shows how the average pressure suddenly increases to 600 bar after roughly 616 ps, while simultaneously the density increases to 1050 g/L (Figure 3b) and the energy drops by 1 kJ/mol (Figure 3c), due to an additional 0.03 hydrogen bonds per molecule (Figure 3f). Indeed we find that the pressure in the direction of the piston (X) is roughly 1, but in the other two directions it is 600−800. In a normal fluid the pressure in the Y and Z would be released through the X dimension, but not in the case when specific ordering is present (like in this artificial case or e.g. in bilayers). To quantify the occurrence of layers we computed the absolute average orientation of the water molecules with respect to the box axes *i*, and from these we computed the mean square deviation from the isotropic value of 0.5[1]

$$\xi = \sum_{i \in x, y, z} \left(\left| \left\langle \frac{\mu_i}{\mu} \right\rangle \right| - 0.5 \right)^2 \qquad (2)$$

where averaging is over all molecules in the computational box. Figure 3d shows ξ as a function of time. Obviously, a gross-net ordering happens simultaneously with the other events, leading to the conclusion that the whole process can be considered to be a liquid−liquid phase transition. Finally, we found that the mobility increases in the new ordered liquid phase. Since we had suspected that some kind of room-temperature freezing was behind the drastic changes of the water properties, we were surprised to see that the diffusion constant actually increases from 7.3 to 9.7 10^5 cm^2 s^{-1}

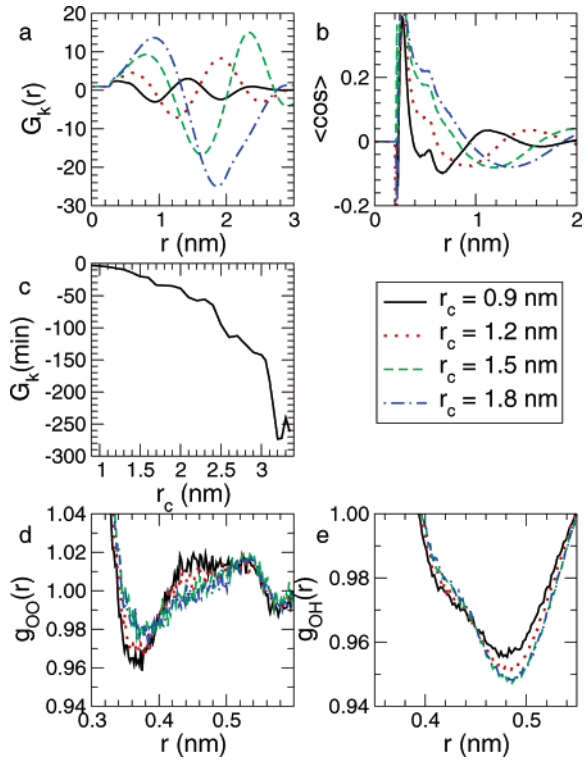


Figure 1. a: The $G_k(r)$ function as a function of the cutoff distance in a TIP3P simulation of 2201 molecules (cutoff r_c indicated in legend), b: ⟨cos⟩ as a function of cutoff in the same set of simulations, c: the depth of the minimum in $G_k(r)$ as a function of cutoff in a series of simulations of 10 648 TIP3P molecules, d: the first minimum in the oxygen−oxygen radial distribution function, and e: the first minimum in the oxygen−hydrogen radial distribution function.

(whereas it is 5.4 for a cutoff of 1.2 nm[17]). On the other hand the increase of diffusion with density and pressure is one of the well-known anomalies of water,[41] and henceforth the high pressure can induce the increased mobility. The artificial water phase can therefore be described as a high-density, high-pressure ordered liquid. There is no special reason anisotropic pressure scaling would facilitate layer formation, other than that in this specific case the width of the box fit the requirement (explained in section 3.4) of being close to an integer times the cutoff.

In Figure 3 we also show what happens when the layered structure is simulated with a short ($r_c = 0.9$ nm) group-based cutoff (from 2000 to 3000 ps). All values fall back quickly to the normal values, showing that layer formation is completely reversible.

3.2. Water Model and Electrostatics Treatment Dependence. To track down how these strange results depend on cutoff treatment and water model, further simulations were performed using the TIP3P and TIP4P models,[2] the TIP5P model,[28] the SPC model,[29] and the SPC/E model.[30] In Figure 4a,f we have plotted again the $G_k(r)$ and the ⟨cos⟩ of the simulations with a 1.8 nm cutoff of the five models. Similarly we have plotted the results from corresponding sets of simulations using a reaction-field (Figure 4b,g), using the PME method (Figure 4c,h), using a switch function (Figure 4d,i), and using a shift function (Figure 4e,j).

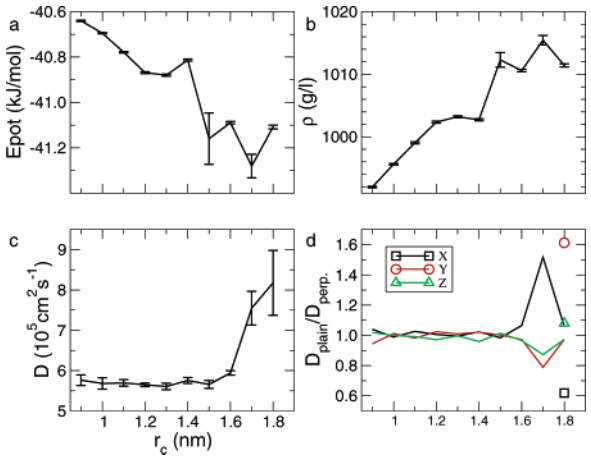


Figure 2. Properties as a function of cutoff in a TIP3P simulation of 2201 molecules using group-based truncation. a: Potential energy, b: density, c: diffusion constant determined by breaking the trajectories in 10 equal pieces of 200 ps, computing the diffusion constant for each molecule separately using the Einstein relation,[23] and taking the average value and standard deviation, and d: ratio of diffusion coefficients in the plane normal to one of the box vectors and along the box vector (for isotropic diffusion the ratio should be 1). The symbols are from the second part of the simulation with anisotropic scaling, i.e., where stable layers are present. Error bars in a and b were determined by a block-averaging procedure.[49]

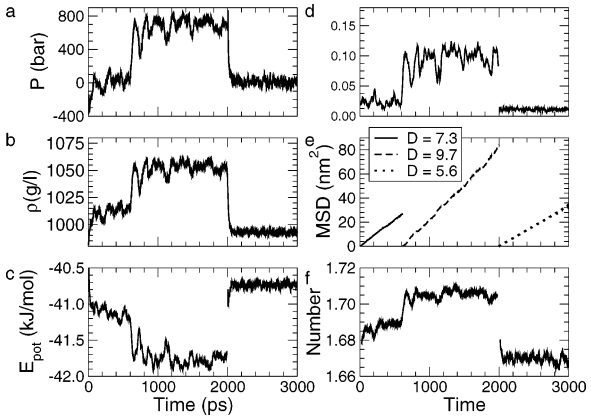


Figure 3. Results from a TIP3P simulation of 2201 molecules where one of the box edges (X) was allowed to fluctuate, while the others were fixed at their initial value of 4.06 nm. The first 2 ns were done with a 1.8 nm cutoff, the third ns was done with a 0.9 nm cutoff, to test reversibility of layer formation. a: Pressure (*bar*), b: density, c: potential energy (kJ/mol), d: ξ (eq 2) and e: mean square displacement computed for three stretches of the trajectory, from 0 to 616 ps (before the phase transition), from 616 to 2000 ps (after), from 2000 to 3000 ps (short cutoff). The resulting diffusion constants (10^5 cm^2 s^{-1}) are indicated, f: the number of hydrogen bonds per molecule. In panels a, c, d, and f a running average over 20 ps is given for clarity.

A number of observations can be made from Figure 4. First, the depression in the $G_k(r)$ function due to the cutoff depends very much on the water model used, but all models are seriously affected. The minimum in Figure 4a is deepest for TIP5P, followed by TIP3P, SPC/E, SPC, and TIP4P.

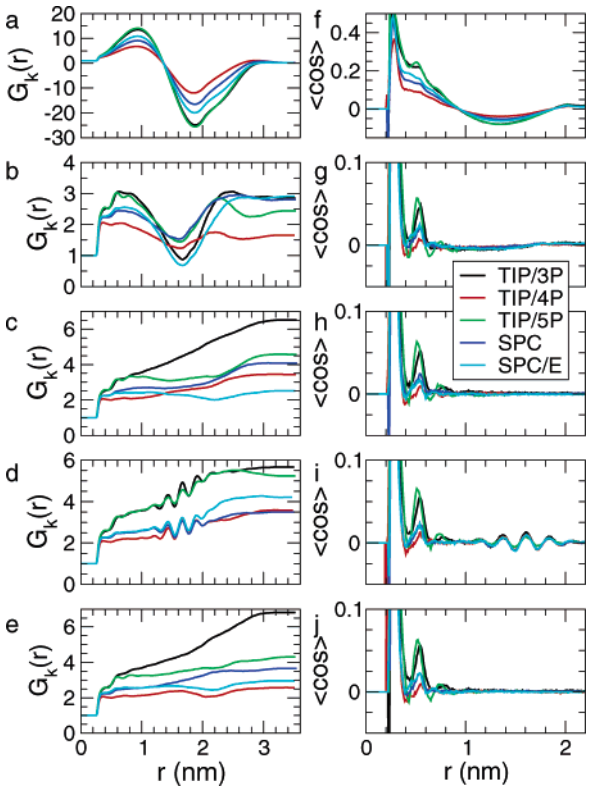


Figure 4. The $G_k(r)$ function for different water models in a cubic box containing 2201 molecules. a: Simulated with 1.8 nm cutoff b: simulated with a reaction field with $\epsilon_{rf} = 78.5$ and a 1.8 nm cutoff, c: simulated with the particle-mesh Ewald technique with a cutoff of 0.9 nm, d: simulated with a switch function and a cutoff of 1.7 nm, and e: simulated with a shift function. The average cosine $\langle\cos\rangle$ of the angle between the dipoles of molecules in a spherical shell at a distance r from a central molecule for f: cutoff, g: reaction-field, h: PME, i: switch function, and j: shift function. Data corresponding to this figure are available as Supporting Information.

Although one would expect the molecular dipole to be decisive about the amount of long-range correlation, in fact we find that SPC/E is much less affected than TIP3P despite having nearly the same dipole, while simultaneously SPC is much less affected than TIP5P. There seems to be a correlation between the depth of the minimum and the magnitude of the molecular quadrupole (Table 1), the only exception being the SPC/E model that is slightly worse than SPC despite having a larger quadrupole. Note that we compare the model quadrupoles (Table 1) to the experimental gas-phase value[42] which need not be the same as the (unknown) liquid-state value.

Second, only the particle-mesh Ewald and the shift function seem to give dependable results that are not affected by the particular choice of cutoff. The shift function, which is a special case of the switch function,[43] seems to alleviate the effects of the cutoff in an efficient way although it obviously will only work correctly when the "real" interaction is zero, i.e., the shift distance should be at least as long as the particular interaction range, which, for water has been estimated to be 1.4−1.5 nm.[39,40] Obviously, ionic interactions cannot be handled faithfully by a shift function. The reason the atom-based switch function gives strange ripples near the cutoff is due to the short switching range which was only 0.2 nm, although this value is commonly used in conjunction with cutoffs in the range of 0.8−1.2 nm. The ripples are reduced somewhat when a longer switching range of 0.4 nm was used (not shown). The alternative of a group-based switch function (Appendix B) is due to the arbitrariness of the group-center definition not attractive in principle, although it probably is devoid of the artifacts (ripples, reduced mobility) we find here. In an attempt to force layer formation using different cutoff treatment, we used anisotropic pressure scaling with 2201 TIP3P molecules. However layering occurred only in combination with the cutoff (Figure 3) and not with any of the other schemes (data not shown).

Table 3. Simulation Results from Simulations of 2201 Molecules Using Five Water Models with Six Different Cutoff Schemes, Corresponding to Figure 4: Cutoff and Reaction Field (RF) Simulations with 1.8 nm Cutoff, Particle-Mesh Ewald (PME) with 0.9 nm Cutoff, Switch and Shift with 1.7 nm Cutoff, and Switch with a 0.2 nm Switching Range (See Methods)[a]

model	cutoff	$D\,10^5$ cm^2 s^{-1}	ϵ_0	ρ (g/L)	E_{pot} (kJ/mol)
expt		2.3	78.5	997	−41.7
ref		66, 67	68	69	70
	cutoff	8.2(0.8)	39(1)	1011.4(0.3)	−41.107(0.008)
	RF	5.96(0.06)	94(3)	982.2(0.2)	−39.968(0.002)
TIP3P	PME	5.76(0.03)	92(4)	970.9(0.2)	−39.882(0.002)
	switch	4.26(0.11)	102(5)	1004.3(0.2)	−40.705(0.003)
	Switch2	5.65(0.15)	103(5)	987.3(0.2)	−40.133(0.002)
	Shift	5.8(0.2)	101(5)	981.5(0.2)	−39.823(0.002)
	Cut-Off	3.88(0.02)	48(1)	1000.8(0.2)	−41.698(0.003)
	RF	3.72(0.02)	54(3)	991.0(0.3)	−41.318(0.003)
TIP4P	PME	3.73(0.02)	49(2)	980.4(0.2)	−41.282(0.002)
	switch	2.65(0.02)	53(2)	1026.5(0.2)	−42.293(0.003)
	switch2	3.53(0.08)	52(2)	998.4(0.2)	−41.528(0.003)
	shift	3.78(0.04)	51(2)	990.3(0.2)	−41.172(0.003)
	cutoff	3.9(0.5)	41(1)	1010.9(0.3)	−41.19(0.01)
	RF	3.02(0.03)	72(3)	981.2(0.2)	−40.282(0.005)
TIP5P	PME	2.95(0.05)	88(7)	969.3(0.2)	−40.232(0.006)
	switch	2.72(0.04)	87(6)	988.6(0.3)	−40.353(0.005)
	switch2	2.75(0.07)	89(6)	983.5(0.3)	−40.521(0.005)
	shift	2.94(0.06)	89(6)	980.4(0.2)	−40.139(0.005)
	cutoff	4.48(0.09)	43(1)	991.2(0.2)	−42.359(0.006)
	RF	4.38(0.05)	65(3)	974.3(0.2)	−41.610(0.003)
SPC	PME	4.29(0.04)	67(3)	963.9(0.2)	−41.535(0.003)
	switch	3.24(0.01)	72(4)	1005.5(0.3)	−42.375(0.006)
	switch2	4.11(0.06)	69(3)	982.3(0.2)	−41.794(0.003)
	shift	4.27(0.11)	62(3)	973.9(0.2)	−41.452(0.003)
	cutoff	2.9(0.2)	43(1)	1014.5(0.3)	−47.55(0.01)
	RF	2.71(0.04)	77(4)	996.0(0.2)	−46.668(0.004)
SPC/E	PME	2.70(0.04)	62(4)	986.5(0.2)	−46.618(0.004)
	switch	2.11(0.02)	74(5)	1027.8(0.3)	−47.414(0.005)
	switch2	2.55(0.01)	76(5)	1003.6(0.2)	−46.873(0.004)
	shift	2.71(0.13)	78(5)	995.7(0.2)	−46.515(0.009)

[a] The error in the dielectric constant was determined by computing the error in the square total dipole moment M^2 of the box using a block averaging procedure[65] and multiplying the relative error in M^2 by the dielectric constant.

A list of important observables from the simulations using different cutoff schemes and water models is given in Table 3. The densities are relatively low due to the fact that we did not use the dispersion correction to the pressure (e.g. Wensink et al. find a density of 994 g/L for TIP4P,[44] while

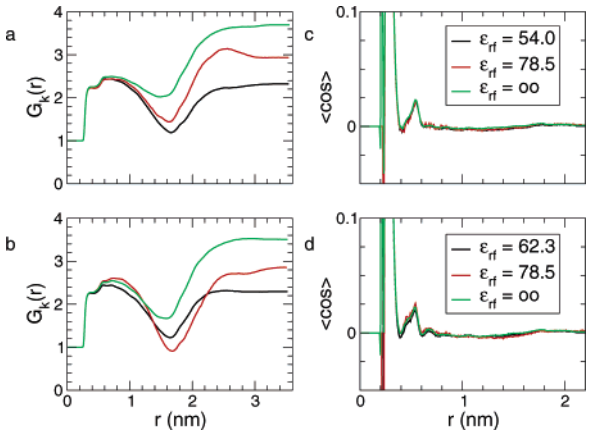


Figure 5. The $G_k(r)$ function for the a: SPC and b: SPC/E water models in a cubic box containing 2201 molecules, simulated with a reaction field with 1.8 nm cutoff and different ϵ_{rf}. The average cosine $\langle\cos\rangle$ of the angle between the dipoles of molecules in a spherical shell at a distance r from a central molecule for c: SPC and d: SPC/E water, under the same conditions as a and b, respectively. Data corresponding to this figure are available as Supporting Information.

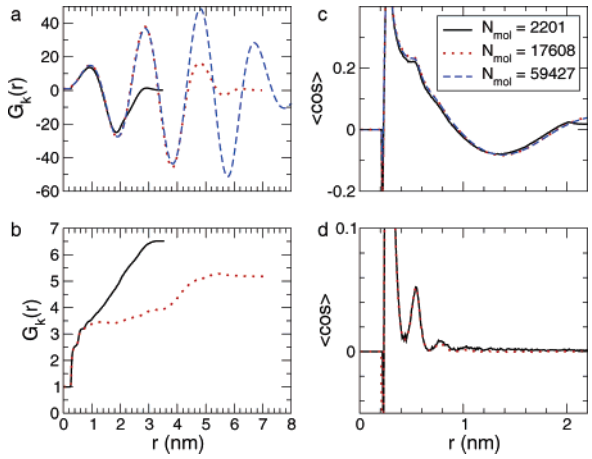


Figure 6. a: The $G_k(r)$ function as a function of system size for TIP3P water simulated with a cutoff of 1.8 nm, b: id. simulated with PME, c: $\langle\cos\rangle$ simulated with a cutoff of 1.8 nm, and d: id. simulated with PME.

Table 4. Simulation Results from Simulations of 2201 Molecules Using Two Water Models with a Reaction Field and Different ϵ_{rf}. Dielectric Constant $\epsilon(0)$ and Depth of Minimum in $G_k(r)$

SPC			SPC/E		
ϵ_{rf}	ϵ_0	G_k(min)	ϵ_{rf}	ϵ_0	G_k(min)
54.0	62(2)	1.19	62.3	67(3)	1.22
65.0	62(2)	1.24	77.0	73(3)	1.22
78.5	65(3)	1.44	78.5	77(4)	0.90
∞	64(4)	2.02	∞	72(4)	1.67

Horn et al. find roughly 988 g/L when using explicit tail corrections to a switched Lennard-Jones potential[45]).

3.3. Effect of Reaction Field Parameters. A comparison highlighting the impact of different ϵ_{rf} when using a reaction-field is given in Figure 5. Smith and Van Gunsteren proposed that it would be better to use a ϵ_{rf} close to the dielectric constant of the simulated liquid.[38] They gave reference values for $\epsilon_{rf} = 54$ for SPC and $\epsilon_{rf} = 62.3$ for SPC/E water. Unfortunately the so-called "self-consistent" ϵ_{rf} are dependent on the cutoff as well, and here we find (Table 3) dielectric constants of 65 (SPC) and 77 (SPC/E). Simulation with both values of the dielectric constant were done for each of the two models. The effect of ϵ_{rf} on the potential is a shift of the minimum of the potential. For $\epsilon_{rf} = \infty$ (conducting boundary conditions), the potential has its minimum exactly at the cutoff, for finite ϵ_{rf} the minimum is shifted to (slightly) larger distances. For $\epsilon_{rf} = \infty$ the minimum in $G_k(r)$ is less pronounced than for finite ϵ_{rf} (Figure 5). The self-consistent values are very comparable to the other values that are not self-consistent; however, all of these affect the $G_k(r)$ function considerably more than $\epsilon_{rf} = \infty$ which hence seems to be the best choice for simulations of water. The dielectric constants computed from the different simulation as well as the depth of the minima in the $G_k(r)$ function are given in Table 4. We conclude that use of the self-consistent ϵ_{rf} should be avoided because it is dependent on cutoff parameters, not transferable to other simulation systems, and because using $\epsilon_{rf} = \infty$ yields considerably less disturbance of the dipole−dipole correlation. As it was shown before that one needs a cutoff of 4.0 nm before reaction field methods yield the same results as Ewald summation,[12] it is questionable whether reaction fields should be used at all.

3.4. Explanation for Layer Formation. Yonetani proposed[1] that the layering effect is in fact due to the cutoff, rather than due to the periodic boundary conditions. To prove this, he used simulations cells that were at most twice the volume of the original. In Figure 6 we have plotted again the $G_k(r)$ and $\langle\cos\rangle$ for cells 8 respectively 27 times the original cell, simulated with a cutoff. In addition, the 8-fold cell was simulated with PME. The size-dependence on $\langle\cos\rangle$ is very small, both when using a cutoff and when using PME, indicating that the relative orientation of the water molecules is not influenced by the size of the system or the periodic boundary conditions. Moreover, this shows that the particle-mesh Ewald method does not impose any artifacts on a water system of this size, which is the typical size for the solvation of a small protein. However, by a careful analysis of the dipole orientation parallel to the box axes (Figure 7) we find that there is a strong periodicity in the dipole along certain axes if the cutoff fits an integer number of times in the box. For a cutoff of 0.9 nm we find four periods in a box of roughly 4 nm, for a cutoff of 1.3 nm we find three periods, and for 1.8 nm we find 2 periods. However for a cutoff of 1.4 nm we find no obvious preference along the box axes, while it seems quite obvious that some orientational preference must exist. This can be explained if the layers are not parallel to the box axes but at an angle. In Figure 8 a schematic representation of such layer structures as observed in our simulations is given. Obviously, the layers are three-dimensional, but some "rules" can be inferred from the structures: the layers are equally thick and the layer thickness corresponds to roughly half the cutoff, or slightly more than so, and finally the number of layers must be even. In the case that the cutoff does not "fit" an integer times in the

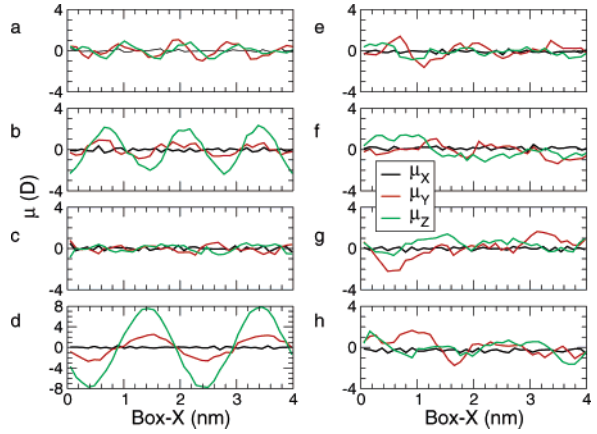


Figure 7. The components of the dipole moment summed over 40 slabs of roughly 0.1 nm along the *X*-axis of the box, for 2201 TIP3P molecules, simulated with different electrostatics treatment. a: Simulated with 0.9 nm cutoff, b: simulated with 1.3 nm cutoff, c: simulated with 1.4 nm cutoff, d: simulated with 1.8 nm cutoff (note different scale on the *Y*-axis), e: simulated with a reaction field with $\epsilon_{rf} = 78.5$ and a 1.8 nm cutoff, f: simulated with the particle-mesh Ewald technique with a cutoff of 0.9 nm, g: simulated with a switch function and a cutoff of 1.7 nm, and h: simulated with a shift function.

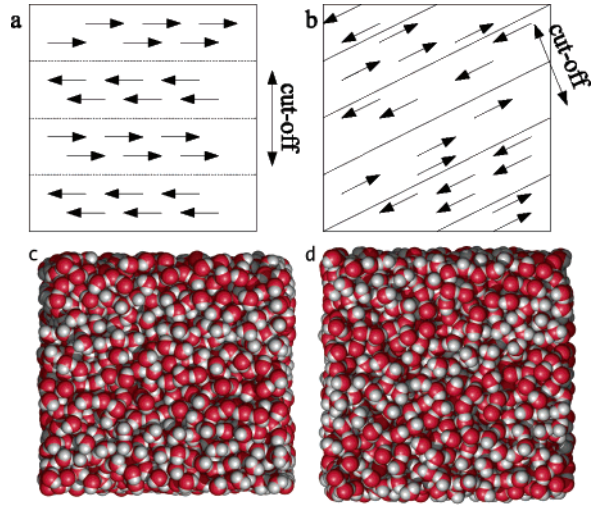


Figure 8. Schematic 2D picture of layer formation, a: four layers parallel to a box plane and b: four layers at an angle. The orientation of the dipoles in each layer as well as the relation between layer thickness and cutoff is indicated; the layers are somewhat thicker than half the cutoff. Snapshots of c: 2201 TIP3P molecules simulated with a cutoff of 1.8 nm (4 layers) and d: 2201 TIP3P molecules simulated with a cutoff of 1.5 nm (4 layers at an angle). Obviously when the box is cut at an angle the layers get thinner, to keep the same volume per slab.

box, the layers can form at an angle α given by

$$\cos \alpha = \frac{N(r_l)}{2l} \tag{3}$$

where N is the number of layers, r_l is the thickness of the layer, and l is the length of the box axis. In Figure 8d we estimate $\alpha \approx 25°$ yielding $r_l = 1.82$ nm, longer than the cutoff $r_c = 1.5$. For a similar snapshot (not shown) from the simulation with $r_c = 1.7$ nm, we find $\alpha \approx 14°$ yielding $r_l = 1.95$ nm. Apparently the layers can be 0.2−0.3 nm thicker than the cutoff. It is important to note that even with a moderate cutoff of 1.5, used in many simulations of biomolecules, significant layer formation takes place. In some cases (mainly in larger systems) we even observed multiple layer systems that existed simultaneously in the simulation, with the result that the value of ξ (eq 2) is low, making detection of layer formation problematic. The simulations performed with reaction field, PME, and switch and shift function do not display any regular pattern in the dipoles (Figure 7). The direct correlation between cutoff induced layer formation and periodic boundary conditions disagrees with Yonetani's conclusion that this is a pure cutoff effect.[1]

3.5. Other Systematic Tests. A number of other simple tests were performed in order to exclude systematic errors.

- A simulation using the Nosé-Hoover thermostat[46,47] and Parrinello-Rahman barostat[48] of 2201 TIP3P molecules with a 1.8 nm cutoff yielded identical $G_k(r)$ to the simulation using the Berendsen thermostat and barostat.[31] The diffusion constant in the Nose-Hoover simulation was larger (8.6 ± 0.2) than the one using Berendsen coupling (8.2 ± 0.8, Table 3) which could be due to systematic reduction of velocities by the Berendsen thermostat, although the difference is small and the error margin quite large.
- Simulations of 2201 TIP3P molecules were done with different pressure coupling constants τ_P in order to check its influence on layer formation and water mobility. For $\tau_P = 0$ (NVT simulation), 0.2, 0.5, and 2.0 ps no detectable difference in ξ (eq 2) could be found, and the $G_k(r)$ were very similar as well (not shown). The diffusion constants were 8.7 (NVT), 8.6 (0.2 ps), 8.2 (0.5 ps), and 8.8 10^5 cm^2 s^{-1} (2.0 ps), respectively. There is no clear trend that can be observed, and hence we conclude that τ_P is not relevant for layer formation.
- The TIP3P simulations using a group-based truncation with $r_c = 1.8$ nm and with the PME algorithm were continued to 10 ns. The $G_k(r)$ functions were virtually identical to those in the shorter simulations (data not shown). In combination with the finding that layer formation is reversible (Figure 3) this shows that our 2 ns simulations are long enough to justify our conclusions.
- All water models were simulated with an atom-based switch function and a switching range of 0.4 nm as well. Although the ripples that are present in Figure 4d have reduced amplitude with the longer switching range (not shown), they do not disappear altogether, implying that the atom-based switch function is not a viable solution for electrostatics treatment.[24] The observables in Table 3 lie between the 0.2 nm switching range and a shift function, as expected.
- A simulation of 2201 TIP3P molecules using double precision arithmetic (rather than the GROMACS default of single precision) yields identical $G_k(r)$ (not shown).

4. Discussion

In this work we have shown that an inappropriate use of simulation parameters can lead to artificial phase transitions.

TIP3P and TIP5P water and to some extent SPC/E and SPC as well were shown to form layers and change density abruptly (Figure 3b). The ordered layers are stabilized by a slightly larger number of hydrogen bonds (Figure 3f), leading to a lower potential energy. In addition, self-diffusion becomes anisotropic, and diffusion is considerably faster in the direction perpendicular to the layers (Figure 2). This artificial stabilization of the layered phase can be expected for all water models when simulated with group-based truncation of the potential; the extent to which it occurs varies between models as discussed below.

Figure 1 shows how the dipole−dipole correlation changes with increasing cutoff. It can be seen that there is a distinct minimum corresponding to anticorrelation (i.e. opposite dipoles[49]), which gets wider with increasing cutoff, while simultaneously the first, positive peak also gets wider. The turning point lies roughly at half the cutoff distance (0.6 nm for 1.2 nm cutoff, 0.84 for 1.5 nm, and 0.94 for 1.8 nm), and this turning point is unrelated to the real structure of water. This implies that for a given water molecule there is a large number of molecules with similar orientation until half the cutoff and an even larger number of molecules (due to the volume of the shell) with opposite orientation (Figure 8). The reason for this effect is that with a group-based cutoff, each water molecule is in effect at the center of a water cluster surrounded by vacuum, the free energy of one of these clusters is minimized by minimizing its net polarization, and hence the dipole moments become anticorrelated.[50] In combination with periodic boundary conditions this effect can lead to layer formation (Figure 7) which further strengthens the interaction. Figure 1c shows that the effect gets even more significant with longer cutoffs. Yonetani[1] suggested that the effect is entirely due to the cutoff; however, Figure 7 shows that some degree of ordering is always present when using a group-based cutoff, and it is particularly obvious when the box size is an (even) integer times the cutoff length. In extreme cases (Figure 3) periodic boundary conditions can enhance and stabilize the layers (Figure 8), in a similar fashion as an artificial wormlike periodic micelle was found to be induced by periodic boundary conditions.[51]

Mathias and Tavan have convincingly shown that there is no orientational correlation in water beyond 1.5 nm.[40] Our simulations with the particle mesh Ewald method[7,8] (Figure 4h) and with the shift function (Figure 4j) both agree with this observation. Although the shift function is well behaved in our simulations, it can obviously not be used for charged systems, while in simulations of neutral molecules the cutoff should be on the order of the real correlation length (for water 1.5 nm[40]). In addition, an ad-hoc addition of a shift function (in contrast to a switch) to a potential optimized without a shift function will change the relative energies in the potential at short distances (note that e.g. the ENCAD force field was calibrated for use with shifted potentials[43]). It seems therefore that a shift function is not an economic choice for neutral systems, while it is inappropriate for charged systems, despite being considerably more accurate than a cutoff.[52] A group-based switch function (Appendix B) could probably diminish the artifacts shown in Figure 4 and indeed be used without reparametrization of the force field, as the short-range interaction is unmodified. Like the shift function it cannot faithfully simulate ionic interactions. Finally, we have shown here that, at least for systems of 4 nm and larger, the PME method does not influence the orientational correlation (Figure 6) which confirms the findings of Mathias and Tavan.[40]

A question that remains is that of the difference between the water models. It has been shown that a large quadrupole can effectively quench the interactions between dipoles, leading to a reduced dielectric constant.[53−55] This is apparently what happens here: TIP4P has the largest quadrupole of the models tested (and the lowest dielectric constant) and is the least affected by the cutoff problems. In this context one could conclude that TIP4P is the most robust of the models used here. Even though e.g. the density maximum of water cannot be reproduced with TIP4P,[56] while that is possible with TIP5P,[28] it seems that the quadrupole of TIP4P is more realistic (note though, that the TIP4P variant that was optimized for use with Ewald summations, TIP4P-Ew[45] does have a density maximum close to the experimental one). Of the empirical models tested here, TIP4P is the only one with a realistic dimer structure,[57] even though the O−O distance is too short due to the effective charges. In addition, the latest versions of the OPLS force field[58] have been tuned for use with the TIP4P model, and simulations of proteins with the OPLS force field and TIP4P water are now beginning to appear.[59,60]

The question whether there is predictive value in water simulations[61] remains intriguing. Although Brodsky answered his own question with a clear no,[61] Guillot has taken a more constructive position, when comparing water models in a review recently.[62] It is somehow ironic that modeling a real phase transition, like freezing[20] takes enormous amounts of computer time, because it is a rare process, while artificial phase transitions such as the one reported here and by Yonetani[1] happen very fast. Van Gunsteren and Mark have published a series of criteria for the validation of molecular dynamics simulations.[27] An important criterion is that the quality of the result depends on the quality of the interaction function (including force field). From our results it is clear that only the particle-mesh Ewald[7,8] and the shift function yield a correct dipole−dipole correlation,[40] while in principle a group-based switch function should also give reliable results. Since neither a shift function nor a switch function can be used for charged systems, only methods that take the full Coulomb interaction into account remain as an option for biomolecular simulation.[26] A further method that we have not tested in this work is the Lekner summation[13,14] which is reported to give results that are in good agreement with experiments for liquid water[16] and which has been used for biomolecular simulation as well.[15]

Appendices

A. Coulomb Interaction with Reaction Field. The coulomb interaction can be modified for homogeneous systems, by assuming a constant dielectric environment beyond the cutoff r_c with a dielectric constant of ϵ_{rf}. The interaction then reads

$$V_{crf} = f\frac{q_i q_j}{r_{ij}}\left[1 + \frac{\epsilon_{rf} - 1}{2\epsilon_{rf} + 1}\frac{r_{ij}^3}{r_c^3}\right] - f\frac{q_i q_j}{r_c}\frac{3\epsilon_{rf}}{2\epsilon_{rf} + 1} \quad (4)$$

where $f = 1/(4\pi\epsilon_0)$ and ϵ_0 is the vacuum permittivity. The constant expression on the right makes the potential zero at the cutoff r_c. Note that at distances larger than r_c the potential *increases* which is relevant in the case of molecular based cutoffs.

B. Form of the Shift and Switch Functions. There is no fundamental difference between a shift function, which modifies a potential over its whole range ($0 \le r < r_c$), and a switch function, which modifies a potential over part of the range ($r_1 \le r < r_c$), since letting $r_1 = 0$ reduces a switch function to a shift function.[43] Switch or shift functions $S(r)$ can be applied to either the energy function $U(r)$ or the force function $F(r)$. In general a weighting function $W(r, r_1, r_c)$ is introduced:

$$W(r, r_1, r_c) = \begin{cases} 1 & \text{if } r < r_1 \\ \mathrm{S}(r,r_1,r_c) & \text{if } r_1 \le r < r_c \\ 0 & \text{if } r_c \le r \end{cases} \quad (5)$$

The switching function can be applied to atoms or to groups of atoms. In the latter case a definition of a group has to be made (e.g. the center of mass) and the value of the switching function is based on the distance R between group centers, and the value $S(R)$ is applied to all pairs of interactions.[24] Below we compare some switching functions $S(r)$ used in the literature to the one used in this work.

B.1. CHARMm Shift. In CHARMm[63] the Coulomb and Lennard Jones energy terms may be shifted ($r_1 = 0$) using

$$S(r, 0, r_c) = \left(1 - \left(\frac{r}{r_c}\right)^2\right)^2 \quad (6)$$

leading to, e.g., a shifted Coulomb force function

$$F_s(r, 0, r_c) = \frac{1}{r^2} + \frac{2}{r_c^2} - \frac{3r^2}{r_c^4} \quad (7)$$

which has a nonzero first derivative at the cutoff distance. This is, however, not a problem if the interactions are computed based on neutral groups.[43]

B.2. CHARMm Switch. A further option in CHARMm is to use a more involved switching function to the energy

$$S(r, r_1, r_c) = \frac{(r_c - r)^2(r_c + 2r - 3r_1)}{(r_c - r_1)^3} \quad (8)$$

this leads to the following Coulomb force function

$$F_s(r, r_1, r_c) = \frac{(r - r_c)(4r^2 + rr_c - 3r_1 r + r_c^2 - 3r_1 r_c)}{r^2(r_1 - r_c)^3} \quad (9)$$

This force has nonzero first derivatives at both r_1 and r_c. The authors of ref 64 mention problems with the robustness of the algorithms they used due to the discontinuities in the second derivative of eq 8.

B.3. ENCAD Shift. The shift function used for the ENCAD force field[43] is applied to the individual van der Waals energy and the Coulomb energy terms $U(r)$ according to

$$S(r, 0, r_c) = 1 - \frac{U(r_c)}{U(r)} - \frac{r - r_c}{U(r)}\frac{\partial U(r_c)}{\partial r} \quad (10)$$

for the Coulomb interaction this gives the following shift function

$$S(r, 0, r_c) = \left(1 - \left(\frac{r}{r_c}\right)\right)^2 \quad (11)$$

leading to a shifted Coulomb force function

$$F_s(r, 0, r_c) = \frac{1}{r^2} - \frac{1}{r_c^2} \quad (12)$$

which, like the CHARMm function, has a nonzero first derivative at the cutoff r_c.

B.4. OPLS Switch. Since the late 1980s OPLS force fields have been developed with a group-based switch function applied to the energy function (since the potentials are used for Monte Carlo no forces are used). The form is

$$S(r, r_1, r_c) = \frac{(r_c^2 - r^2)}{(r_c^2 - r_1^2)} \quad (13)$$

with r_1 typically $r_c - 0.05$ nm. This gives the following switched Coulomb force function:

$$F_s(r, r_1, r_c) = \frac{(r_c^2 - r^2)}{r^2(r_c^2 - r_1^2)} \quad (14)$$

The TIP5P model was developed with this switch function, but not the TIP3P and TIP4P models (W. L. Jorgensen, private communication), which, like SPC and SPC/E, were parametrized with group based truncation.

B.5. Ohmine Switch. A further switch function is used by Ohmine et al.[21,20] to multiply the energy with

$$S(r, r_1, r_c) = \frac{(r - r_c)^3[10(r - r_1)^2 - 5(r - r_1)(r - r_c) + (r - r_c)^2]}{(r_1 - r_c)^5} \quad (15)$$

in practice $r_1 = r_c - 0.2$ nm. This switch function results in a Coulomb force function that has no discontinuities in the first and second derivatives.

B.6. GROMACS Switch. The GROMACS switch[4−6] used in this work is applied to the force function $F(r)$ and is given by

$$S(r, r_1, r_c) = \frac{1 + A(r - r_1)^2 + B(r - r_1)^3}{r^{-(\alpha+1)}} \quad (16)$$

where α is the power of the interaction (1 for Coulomb, 6 and 12 for dispersion and repulsion, respectively). The constants A and B follow from the conditions that the function should be smooth at r_1 and r_c, hence

$$A = -\frac{(\alpha+4)r_c - (\alpha+1)r_1}{r_c^{\alpha+2}(r_c - r_1)^2}$$

$$B = \frac{(\alpha+3)r_c - (\alpha+1)r_1}{r_c^{\alpha+2}(r_c - r_1)^3} \quad (17)$$

Thus the total force function is

$$F_s(r, r_1, r_c) = \frac{1}{r^{\alpha+1}} + A(r - r_1)^2 + B(r - r_1)^3 \quad (18)$$

When $r_1 = 0$, the modified Coulomb force function is

$$F_s(r, r_1, r_c) = \frac{1}{r^2} - \frac{5r^2}{r_c^4} + \frac{4r^3}{r_c^5} \quad (19)$$

Like the Ohmine switch, this function has a smooth force at r_c and at r_1.

Supporting Information Available: Data corresponding to Figures 4 and 5 This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Yonetani, Y. *Chem. Phys. Lett.* **2005**, *406*, 49−53.

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

(3) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham III, T. E.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. *Comput. Phys. Comm.* **1995**, *91*, 1−41.

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

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

(6) van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *J. Comput. Chem.* **2005**, *26*, 1701−1718.

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

(8) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577−8592.

(9) Hockney, R. W.; Eastwood, J. W. *Computer simulation using particles;* McGraw-Hill: New York, 1981.

(10) Hünenberger, P. H. *J. Chem. Phys.* **2002**, *116*, 6880−6897.

(11) Greengard, L.; Rokhlin, V. *J. Comput. Phys.* **1987**, *73*, 325−348.

(12) Mathias, G.; Egwolf, B.; Nonella, M.; Tavan, P. *J. Chem. Phys.* **2003**, *118*, 10847−10860.

(13) Lekner, J. *Physica A* **1989**, *157*, 826−838.

(14) Lekner, J. *Physica A* **1991**, *176*, 485−498.

(15) Juffer, A. H.; Shepherd, C. M.; Vogel, H. J. *J. Chem. Phys.* **2001**, *114*, 1892−1905.

(16) English, N. J. *BTmph* **2005**, *103*, 1945−1960.

(17) van der Spoel, D.; van Maaren, P. J.; Berendsen, H. J. C. *J. Chem. Phys.* **1998**, *108*, 10220−10230.

(18) Slovák, J.; Tanaka, H. *BTjcp* **2005**, *122*, 204512.

(19) Zangi, R.; Mark, A. E. *Phys. Rev. Lett.* **2003**, *91*, 025502.

(20) Matsumoto, M.; Saito, S.; Ohmine, I. *Nature* **2002**, *416*, 409−413.

(21) Ohmine, I.; Tanaka, H.; Wolynes, P. G. *J. Chem. Phys.* **1988**, *89*, 5852−5860.

(22) Yamada, M.; Mossa, S.; Stanley, H. E.; Sciortino, F. *Phys. Rev. Lett.* **2002**, *88*, 195701.

(23) Allen, M. P.; Tildesley, D. J. *Computer Simulations of Liquids;* Oxford Science Publications: Oxford, 1987.

(24) Leach, A. R. *Molecular modelling: principles and applications;* Addison-Wesley Longman: Harlow, U.K., 1996.

(25) Koga, K.; Tanaka, H.; Zeng, X. C. *Nature* **2000**, *408*, 564−567.

(26) Hermans, J. *Proteins: Struct., Funct., Genet.* **1997**, *27*, i.

(27) van Gunsteren, W. F.; Mark, A. E. *J. Chem. Phys.* **1998**, *108*, 6109−6116.

(28) Mahoney, M. W.; Jorgensen, W. L. *J. Chem. Phys.* **2000**, *112*, 8910−8922.

(29) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. In *Intermolecular Forces*; Pullman, B., Ed.; D. Reidel Publishing Company: Dordrecht, 1981; pp 331−342.

(30) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269−6271.

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

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

(33) Harvey, S. C.; Tan, R. K. Z.; Cheatham, T. E. *J. Comput. Chem.* **1998**, *19*, 726−740.

(34) Watts, R. O. *Mol. Phys.* **1974**, *28*, 1069−1083.

(35) van Gunsteren, W. F.; Berendsen, H. J. C.; Rullmann, J. A. C. *Discuss. Faraday Soc.* **1978**, *66*, 58−70.

(36) Neumann, M.; Steinhauser, O. *Mol. Phys.* **1980**, *39*, 437−454.

(37) Neumann, M. *Mol. Phys.* **1983**, *50*, 841−858.

(38) Smith, P. E.; van Gunsteren, W. F. *J. Chem. Phys.* **1994**, *100*, 3169−3174.

(39) Chipot, C.; Milot, C.; Maigret, B.; Kollman, P. A. *J. Chem. Phys.* **1994**, *101*, 7953−7962.

(40) Mathias, G.; Tavan, P. *J. Chem. Phys.* **2004**, *120*, 4393−4403.

(41) Ludwig, R. *Angew. Chem., Int. Ed.* **2001**, *40*, 1808−1827.

(42) Buckingham, A. D. *Quart. Rev. (London)* **1959**, *13*, 183−214.

(43) Levitt, M.; Hirshberg, M.; Sharon, R.; Daggett, V. *Comput. Phys. Comm.* **1995**, *91*, 215−231.

(44) Wensink, E. J.; Hoffmann, A. C.; van Maaren, P. J.; van der Spoel, D. *J. Chem. Phys.* **2003**, *119*, 7308−7317.

(45) Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. *J. Chem. Phys.* **2004**, *120*, 9665−9678.

(46) Nosé, S. *Mol. Phys.* **1984**, *52*, 255−268.

(47) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695−1697.

(48) Parrinello, M.; Rahman, A. *J. Appl. Phys.* **1981**, *52*, 7182−7190.

(49) Hess, B. *J. Chem. Phys.* **2002**, *116*, 209−217.

(50) Straatsma, T. P.; Berendsen, H. J. C. *J. Chem. Phys.* **1988**, *89*, 5876−5886.

(51) Marrink, S. J.; Tieleman, D. P.; Mark, A. E. *J. Phys. Chem. B* **2000**, *104*, 12165−12173.

(52) Beck, D. A. C.; Armen, R. S.; Daggett, V. *Biochemistry* **2005**, *44*, 609−616.

(53) Carnie, S. L.; Patey, G. N. *Mol. Phys.* **1982**, *47*, 1129.

(54) Fries, P. H.; Richardi, J.; Krienke, H. *Mol. Phys.* **1997**, *90*, 841−854.

(55) Rick, S. W. *J. Chem. Phys.* **2004**, *120*, 6085−6093.

(56) Jorgensen, W. L.; Jenson, C. *J. Comput. Chem.* **1998**, *19*, 1179−1186.

(57) van Maaren, P. J.; van der Spoel, D. *J. Phys. Chem. B* **2001**, *105*, 2618−2626.

(58) Jorgensen, W. L. Wiley: New York, 1998; Vol. 3, chapter OPLS Force Fields, pp 1986−1989.

(59) van der Spoel, D.; Lindahl, E. *J. Phys. Chem. B* **2003**, *117*, 11178−11187.

(60) Seibert, M.; Patriksson, A.; Hess, B.; van der Spoel, D. **2005**, in press.

(61) Brodsky, A. *Chem. Phys. Lett.* **1996**, *261*, 563−568.

(62) Guillot, B. *J. Mol. Liq.* **2002**, *101*, 219−260.

(63) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.

(64) Leslie, L. J.; Berne, B. J. *J. Phys. Chem. A* **1997**, *107*, 4350−4357.

(65) Hess, B.; Saint-Martin, H.; Berendsen, H. J. C. *J. Chem. Phys.* **2002**, *116*, 9602−9610.

(66) Jonas, J.; DeFries, T.; Wilbur, D. J. *J. Chem. Phys.* **1976**, *65*, 582−588.

(67) Price, W. S.; Ide, H.; Arata, Y. *J. Phys. Chem. A* **1999**, *103*, 448−450.

(68) Weast, R. C. *Handbook of Chemistry and Physics;* CRC Press: Cleveland, OH, 1977.

(69) Schmidt, E. *Properties of Water and Steam in SI-Units;* Springer-Verlag: Berlin, 1969.

(70) Postma, J. P. M. A Molecular Dynamics Study of Water, Ph.D. Thesis, University of Groningen, 1985.

CT0502256

JCTC Journal of Chemical Theory and Computation

Preaveraged Hydrodynamic Interaction Revisited via Boundary Element Computations

Sergio R. Aragon* and David K. Hahn

Department of Chemistry and Biochemistry, San Francisco State University, 1600 Holloway Avenue, San Francisco, California 94132-4163

Received June 20, 2005

Abstract: The effect of preaveraging the Oseen tensor to yield a scalar approximation is examined for transport problems of rigid objects with stick boundary conditions using new very high accuracy computational codes. Nearly exact computations are compared to analytical results and preaveraged results for spheroids and, similarly, for a set of three globular proteins. In agreement with previous work, we find that the error in translational diffusion is less than 1%. However, in the case of rotational diffusion and intrinsic viscosity, the error is sensitively dependent on shape. In the case of the axial component of the rotational diffusion, the error is about −34% independent of shape, but for the perpendicular component, the error starts at −30% (sphere) and decreases as the axial ratio increases and then yields a similar but positive error. For the instrinsic viscosity, the errors are around 10% near spherical and decrease toward the needle or disk shape. For the globular proteins, the errors are similar to those found for the ellipsoids near the spherical shape. The calculations show that preaveraging is acceptable only for translational diffusion of rigid objects.

I. Introduction

It is well-known that the hydrodynamic interaction between two spheres is a series expansion whose first term is the Oseen[1] tensor, and to second order in the distance between the spheres, the interaction can be variationally represented by the Rotne−Prager[2] tensor. The tensorial nature of the hydrodynamic interaction, HI, requires more extensive computations when doing bead modeling dynamics of polymers and often the interaction is orientationally averaged to save computation time,[3] but this may not always be justified. When such orientational averaging is performed, the Oseen tensor becomes proportional to the Green function for electrostatic problems. Several authors have taken advantage of this and have developed extensive formulas that connect, in an approximate manner, the transport properties of rigid bodies with electrostatic properties of these bodies. In particular, the capacitance of an arbitrarily shaped conductor has been related to the translational diffusion coefficient,[4] while the polarizability has been related to the intrinsic viscosity[5] and the normal component of the polarizability has been related to the rotational diffusion coefficient of the body.[6] Zhou[6] has further developed these formulas and has applied them to the computation of the transport properties of proteins by the use of a boundary element method. As we show below, such computations can be very inaccurate, requiring the use of empirical correction factors to gain agreement with experiment.

Studies based on interacting bead hydrodynamics by Garcia de la Torre and co-workers[7] have shown that the translational diffusion coefficient is not sensitive to the tensorial part of the HI tensors so that the preaveraged approximation is expected to work well. On the other hand, accurate enough hydrodynamic computations of rigid body transport have not been available until recently in order to quantitatively determine the error in this approximation, in particular for the rotational diffusion and intrinsic viscosity. In this work, we perform hydrodynamic computations of high accuracy to address this issue, restricted to stick boundary conditions.

II. Theory

In a recent paper, Aragon[8] has described a very accurate implementation of the method introduced by Youngren and Acrivos[9] for the boundary element method of solution of

10.1021/ct050158f CCC: $33.50
Published on Web 11/09/2005

the exact integral equation formulation of the resistance problem. This integral equation has also been used by Allison[10,11] to study the electrophoretic mobility of proteins and DNA. For stick boundary conditions, the equation yields the velocity field at the surface of a body moving in a solvent which is quiescent at infinity. The unknown quantities are the surface stresses (force/area), $\mathbf{f}(\mathbf{x})$. The equation is

$$\mathbf{v}(\mathbf{y}) = \int_{sp} \ddot{\mathbf{T}}(\mathbf{x},\mathbf{y})\cdot\mathbf{f}(\mathbf{x})\, dS_x \tag{1}$$

This is the exact solution to the Stokes equation of hydrodynamics under stick boundary conditions, and it applies to the case of small Reynold's number flow. The kernel of the equation is the Oseen tensor given by

$$\ddot{\mathbf{T}}(\mathbf{x},\mathbf{y}) = \frac{1}{8\pi\eta|\mathbf{x}-\mathbf{y}|}\left[\ddot{\mathbf{I}} + \frac{(\mathbf{x}-\mathbf{y})(\mathbf{x}-\mathbf{y})}{|\mathbf{x}-\mathbf{y}|^2}\right] \tag{2}$$

The integral eq 1 is solved numerically by replacing the surface with a collection of N triangles that smoothly tile the molecular surface. Then we can write

$$S = \sum_{j=1}^{N} \Delta_j \tag{3}$$

We place the coordinate $\mathbf{x}_j$ at the center of the small triangle Δ_j and take the surface stress force $\mathbf{f}(\mathbf{x})$ to be a constant over the entire patch area. This is the basic approximation: it is clear that it will become a better and better approximation as the triangle is made small. Thus, an extrapolation to zero size triangle leads to a very precise value for the transport properties. With this approximation, eq 1 becomes a set of $3N$ equations for $3N$ unknowns $\mathbf{f}(\mathbf{x})$

$$\mathbf{v}(\mathbf{y}_k) = \sum_{j=1}^{N} \ddot{\mathbf{G}}_{kj}\cdot\mathbf{f}_j \tag{4}$$

The centerpiece of this set of equations is a set of N completely known 3×3 matrices of coefficients that contain all geometric information, the integrals of the Oseen tensor over a surface patch

$$\ddot{\mathbf{G}}_{kj} = \int_{\Delta_j} \ddot{\mathbf{T}}(\mathbf{x},\mathbf{y}_k)\, dS_x \tag{5}$$

The set of $3N$ equations can be written all at once

$$\begin{bmatrix} \mathbf{v}_1 \\ .. \\ .. \\ \mathbf{v}_N \end{bmatrix}_{3Nx1} = \begin{bmatrix} \ddot{\mathbf{G}}_{11} & .. & .. & \ddot{\mathbf{G}}_{1N} \\ .. & .. & .. & .. \\ .. & .. & .. & .. \\ \ddot{\mathbf{G}}_{N1} & .. & .. & \ddot{\mathbf{G}}_{NN} \end{bmatrix}_{3Nx3N} \begin{bmatrix} \mathbf{f}_1 \\ .. \\ .. \\ \mathbf{f}_N \end{bmatrix}_{3Nx1} \tag{6}$$

from which the unknown surface stress forces can be readily obtained by matrix inversion of the $3N\times3N$ super matrix $\ddot{\mathbf{G}}$

$$[\mathbf{f}]_{3N\times1} = [\ddot{\mathbf{G}}]_{3N\times3N}{}^{-1}[\mathbf{v}]_{3N\times1} \tag{7}$$

The total force and torque on the body can be computed from the surface stress forces, and these are directly related to the friction tensors ($\ddot{\mathbf{K}}$) of the body

$$\mathbf{F} = \sum_{j=1}^{N} \mathbf{f}_j(\mathbf{x})\Delta_j = -\ddot{\mathbf{K}}_{tt}\cdot\mathbf{v}_p - \ddot{\mathbf{K}}_{tr}\cdot\omega_p \tag{8}$$

$$\mathbf{T} = \sum_{j=1}^{N} \mathbf{x}_p \times \mathbf{f}_j(\mathbf{x})\Delta_j = -\ddot{\mathbf{K}}_{rt}\cdot\mathbf{v}_p - \ddot{\mathbf{K}}_{rr}\cdot\omega_p \tag{9}$$

The particle can be assumed to have specific translation velocity $\mathbf{v}_p$ and angular velocity ω_p (for example $\omega_p = 0$ and $\mathbf{v}_p = (v_x,0,0)$) to solve the above equations. Thus, 6 calculations suffice to determine all components of the friction tensors. The friction tensors form part of a larger 6×6 tensor that contains information about the pure translational friction (*tt*), the pure rotational friction (*rr*), and the coupling that may exist between these (*rt* and *tr*). There are actually only 3 independent friction tensors because the $\ddot{\mathbf{K}}_{tr}$ tensor is the transpose of the $\ddot{\mathbf{K}}_{rt}$ tensor. The diffusion tensors are finally obtained from the friction tensors by an easy 3×3 matrix inversion

$$\ddot{\mathbf{D}}_{tt} = kT[\ddot{\mathbf{K}}_{tt} - \ddot{\mathbf{K}}_{tr}\cdot\ddot{\mathbf{K}}_{rr}{}^{-1}\cdot\ddot{\mathbf{K}}_{rt}]^{-1} \tag{10}$$

$$\ddot{\mathbf{D}}_{rr} = kT[\ddot{\mathbf{K}}_{rr} - \ddot{\mathbf{K}}_{tr}\cdot\ddot{\mathbf{K}}_{tt}{}^{-1}\cdot\ddot{\mathbf{K}}_{rt}]^{-1} \tag{11}$$

In the final step, the tensors are transformed to the "Center of Diffusion". The details of the procedure, the regularization method, the extrapolation to an infinite number of patches, and the Fortran program, BEST, that accomplished this are described in Aragon.[8]

If one preaverages the HI, then the effective quantity that appears in eq 1 is simply

$$T(\mathbf{x},\mathbf{y}) = \frac{1}{6\pi\eta|\mathbf{x}-\mathbf{y}|} \tag{12}$$

Thus, if one turns off the tensor pieces in the BEST computation (the dyadic following the unit matrix in eq 2), for translation and rotation, one must simply multiply the computed quantity by 4/3. On the other hand, the intrinsic viscosity is proportional to moments of the surface stress forces, and the scaling factor for this case is the inverse, 3/4. This is done automatically in the BEST program.

A comment on why the preaveraging approximation is expected to be a good approximation is in order. Hubbard and Douglas[4] have made the argument that the rotational dynamic motions of the Brownian particle can be considered to perform an operation of angular averaging on the surface stress forces $\mathbf{f}$. From this they have derived the existence of a spherically symmetric stress potential which describes the average flow field around the body and a corresponding relation between the friction coefficient and the electrostatic capacitance. From this relation they proceed to show that preaveraging the hydrodynamic tensor yields estimates that are only a few percent off from known analytical results for a variety of shapes. A crucial factor that allows this to work is that the friction force is a simple integral over the surface stresses over the body. When one considers the intrinsic viscosity or rotational diffusion, on the other hand, two important differences arise. In rotational diffusion, one is interested in the total torque on the body, and now it is a higher moment of the surface stress force that needs to be

Table 1. Translational Diffusion Coefficients of Ellipsoids of Revolution (1/Å)

	prolate and sphere				oblate			
p, $1/p$	exact	BEST	PAV	% err.	exact	BEST	PAV	% err
1	1.3333	1.3333	1.3312	−0.16				
4	0.7104	0.7104	0.7091	−0.19	0.45378	0.45384	0.45282	−0.23
8	0.4651	0.4652	0.4643	−0.19	0.24282	0.24283	0.24228	−0.22
30	0.1821	0.1820	0.1816	−0.22	0.068369	0.068406	0.068031	−0.55

Table 2. Rotational Diffusion Coefficients of Ellipsoids of Revolution (1/Å^3)

	prolate and sphere							
	$D_r^{\perp}$				$D_r^{\parallel}$			
$1/p$	exact	BEST	PAV	% err	exact	BEST	PAV	% err
1	1.0000	1.0002	0.661	−34				
4	7.362×10^{-2}	7.363×10^{-2}	6.452×10^{-2}	−12	3.467×10^{-1}	3.467×10^{-1}	2.281×10^{-1}	−34
8	1.330×10^{-2}	1.330×10^{-2}	1.339×10^{-2}	0.65	1.822×10^{-1}	1.822×10^{-1}	1.1995×10^{-1}	−34
30	3.993×10^{-4}	3.993×10^{-4}	4.595×10^{-4}	15	4.983×10^{-2}	4.956×10^{-2}	3.452×10^{-2}	−30
	oblate							
	$D_r^{\perp}$				$D_r^{\parallel}$			
p	exact	BEST	PAV	% err	exact	BEST	PAV	% err
4	3.391×10^{-2}	3.393×10^{-2}	3.040×10^{-2}	−10	2.778×10^{-2}	2.777×10^{-2}	1.840×10^{-2}	−34
8	4.502×10^{-3}	4.505×10^{-3}	4.771×10^{-3}	5.9	3.964×10^{-3}	3.962×10^{-3}	2.624×10^{-3}	−34
30	8.712×10^{-5}	8.721×10^{-5}	10.83×10^{-5}	24	8.370×10^{-5}	8.372×10^{-5}	5.529×10^{-5}	−34

integrated over the surface of the body ($\mathbf{r} \times \mathbf{f}$); while in the intrinsic viscosity, even more complex higher moments arise. In addition, for rotational diffusion, there are no other dynamics that can serve as the heuristic physical averaging process, and the components of the friction tensor can be quite different for different rotational motions. In the case of translational friction, all the components of the friction tensor are very similar, even when the shape is very anisotropic. A representation by an angular average is therefore expected to work. Furthermore, for the computation of the intrinsic viscosity, we do assume that the rotational Brownian motion does provide an orientational average; however, the more complex moments of the surface stress that enter into the computation provide significant differences between a computation that averages the hydrodynamic interactions at the outset and one that adds the average effect of the higher moments. This, however, probably explains the fact, as shown below, that the error of preaveraging in the intrinsic viscosity is smaller than that for rotational diffusion.

III. Results and Discussion

To investigate the error in preaveraged approximation, we performed computations on ellipsoids of revolution for which analytical formulas exist[12] and for a set of three proteins using our program BEST. We discuss the ellipsoid results first.

A. Ellipsoids. Table 1 shows the values obtained for the average translational diffusion coefficient (1/3 Tr D_t) as a function of axial ratio for both prolate and oblate ellipsoids with three methods: the analytic formulas, the accurate BE solution, and the approximate preaveraged solution. Note that the values in Tables 1 and 2 are given in internal BEST units, 1/A and 1/A^3, respectively, and that a factor of $kT/(8\pi\eta)$ has been taken out for convenience.

The analytic formula for the translational diffusion $\mathbf{D}[p]$ is a function of the axial ratio $p = b/a$, where a is the semiaxis of revolution. Omitting the factor of $kT/(8\pi\eta)$, we have

$$\mathbf{D}_{tt}[p] = 4G[p]/3a \tag{13}$$

where

$$G[p] = \mathrm{Log}\left[\frac{1+\sqrt{1-p^2}}{p}\right]/\sqrt{1-p^2}, \quad \text{for prolate ellipsoids, } p<1 \tag{14}$$

$$G[p] = \mathrm{ArcTan}[\sqrt{p^2-1}]/\sqrt{p^2-1}, \quad \text{for oblate ellipsoids, } p>1 \tag{15}$$

Table 1 shows three computations, including the percent error of the preaveraged approximation (PAV) compared to the analytical formulas. The table demonstrates that the full HI done in BEST is indeed very accurate and that the error in the preaveraged approximation is insignificant. As noted by Douglas and Garbozcsi,[5] for the case of the ellipsoids, the PAV approximation actually yields the exact value. The small discrepancy we observe arises as a small systematic error due to curvature in the extrapolation of the properties to an infinite number of triangles. The full HI computation (BEST) is extremely linear, providing better accuracy. As previously observed in the literature, there is no harm in preaveraging the Oseen tensor for translational diffusion of rigid bodies. In Aragon,[8] it is shown that the full translational

Table 3. Viscosity Factors of Ellipsoids of Revolution

	prolate and sphere				oblate			
p, $1/p$	exact	BEST	PAV	% err	exact	BEST	PAV	% err
1	2.5000	2.4997	2.2611	−9.5				
4	4.6633	4.6626	4.4153	−5.3	4.0593	4.0578	3.7524	−7.5
8	10.103	10.099	9.8616	−2.4	6.7002	6.6985	6.2520	−6.7
30	74.505	74.515	74.751	0.32	21.585	21.564	20.464	−5.1

diffusion tensor is computed in exact agreement with the analytic formulas by BEST.

In Table 2, we show the data similarly arranged for the two eigenvalues of the rotational diffusion tensor. In experiments, the end-over-end rotation (or perpendicular component) is the quantity that is typically observable. The analytic formulas are

$$Drr_{\parallel} = \frac{1}{a^3 p^2} \frac{3(1 - p^2 G[p])}{2(1 - p^2)} \tag{16}$$

$$Drr_{\perp} = \frac{1}{a^3} \frac{3(G[p](2 - p^2) - 1)}{2(1 - p^4)} \tag{17}$$

Table 2 shows once more that the full HI case done in BEST is very accurate and that the error of the preaveraged approximation is quite large, on the order of 30% for either shape in the case of the axial rotation (parallel component), independent of the axial ratio. The error in the perpendicular component is extremely sensitive to the axial ratio. The preaveraged approximation underestimates the tensor by around 30% at small axial ratios and overestimates by a similar amount for an increasing axial ratio. The error is zero for a specific axial ratio dependent on the shape. Preaveraging the Oseen tensor is not appropriate for rotational diffusion. Significant empirical corrections are required in this case, as seen in the work of Zhou.[6]

In Table 3, we show the data for the dimensionless viscosity factor, Σ, of ellipsoids compared to the formula of Simha,[13] assuming negligible orientation of the ellipsoid in the flow field. The intrinsic viscosity is proportional to the viscosity factor, and it can be written in terms of the particle density: $[\eta] = \Sigma/\rho$. The Simha formula is

$$\Sigma = \frac{2(1 - p^2)^2}{15p^2}\left(\frac{3(1 - 2p^2 + p^2 G[p])}{(1 + (p^2 - 2)G[p])(3p^2 G[p] - 2p^2 - 1)} + \frac{3p^2 G[p](5p^2 + 8) - 41p^2 + 2}{(G[p](p^2 + 2) - 3)(3p^4 G[p] - 5p^2 + 2)}\right) \tag{18}$$

Note that the formula given by Richards[18] is incorrect.

Table 3 shows that the intrinsic viscosity is again accurately computed with the full tensor, while the preaveraged approximation makes errors of up to 10%, with significant dependence on axial ratio. The preaveraging yields no error for a prolate ellipsoid of axial ratio near 30, but the error will increase again beyond that. The axial ratio for which this happens in the oblate case is larger. Since the intrinsic viscosity is sensitive only to shape and not size, such errors can lead to a misrepresentation of the shape of the object that could be deduced from these values. In the case of large axial ratios, when the values are used to interpret data on a homologous series with fixed molecular thickness, the shape dependence will yield an overestimate of the molecular length. The errors noted here are comparable to those quoted by Douglas and Garboczi.[5]

Table 4. Polyhedron Transport Properties

	X_t[a]		X_r[a]		ξ	
solid	BEST	PAV	BEST	PAV	BEST	PAV
tetrahedron	0.8229	0.8502	0.4778	0.3940	4.210	3.798
cube	0.922	0.937	0.751	0.544	3.10	2.75
octahedron	0.9318	0.9456	0.7726	0.5583	3.016	2.679
dodecahedron	0.9733	0.9791	0.9137	0.6236	2.691	2.397
icosahedron	0.9808	0.9846	0.9356	0.6344	2.636	2.358

[a] $X_q = D_q/D_q^0$, where D_q^0 is the diffusion coefficient of a same-volume sphere. "t" represents translation, "r" represents rotation, and ξ is the viscosity factor.

B. Polyhedra. Polyhedra can be used to study the dependence of the error on account of the presence of sharp corners in a body.[14] Are the errors in the preaveraged translational friction coefficient significant, in particular, for a tetrahedral shape? To test whether the preaveraged approximation has significant errors dependent on shape, we have computed the viscosity factor, the translational, and the rotational friction coefficients for the entire series of Platonic solids. The data are shown in Table 4.

The data show that there is a shape dependent effect for the translational friction, X_t, but it is small. The largest error does indeed occur for the object that has the sharpest corners, the tetrahedron, but even there it is only 3.3%. The error tends to disappear as the surface tends toward a smooth shape and is less than 1% for the last two members of the series.

The rotational friction coefficient has large errors that range from 21% for the tetrahedron to 47% for the icosahedron. There is significant shape dependence, with the shapes closest to spherical having the largest error, and the size of the errors is comparable to what we find for ellipsoids of revolution. The viscosity factor is uniformly underestimated by 11% with no significant shape dependence.

C. Proteins. To evaluate the effect of the preaveraged HI for irregular shapes, we discuss data obtained for a set of three globular proteins, lysozyme, myoglobin, and human serum albumin (input crystal structures obtained from the Brookhaven database). Figure 1 shows a typical triangulation of the hydrated surface of myoglobin. In Table 4 we show the equivalent data discussed above for the ellipsoids. In this case, we report the average of the rotational diffusion tensor to compare against the accurate computations of BEST. In addition, we have used a water hydration thickness of 1.1 A

Table 5. Hydrodynamic Properties of Proteins[a]

	D_t (10^{-7}cm^2/s)				D_r (10^7/s)				[η] (cm^3/g)			
protein	exp	BEST	PAV	% err	exp	BEST	PAV	% err	exp	BEST	PAV	% err
lysozyme	10.9	11.0	11.1	0.91	2.04	2.09	1.49	−29	3.00	3.22	2.90	−9.9
myoglobin	10.2	10.4	10.3	−0.96	1.46	1.67	1.20	−28	3.25	3.37	2.99	−11
albumin	6.15	6.17	6.23	0.97	0.357	0.349	0.260	−26	3.9	3.92	3.58	−8.7

[a] Lysozyme (2CDS, hen); myoglobin (1MBO, sperm whale); albumin (1AO6, human serum).

Table 6. Protein Translation Friction Coefficients (10^{-11} kg/s)[a]

protein	PDB	f_d	f_s	BEST	PAV	f_k
ferredoxin	1FCA		2.62	2.66	2.63	2.63
ribonuclease S	2RNS	3.79	3.80	3.67	3.63	3.62
lysozyme	2CDS	3.61	3.71	3.68	3.64	3.59
trypsin	1TPO	4.18	4.36	4.25	4.21	4.18
subtilisin BPN′	1SBT	4.48	4.48	4.48	4.43	4.38
carboxypeptidase A	1M4L		4.91	4.86	4.80	4.68
thermolysin	2TLX		4.57	4.97	4.92	4.77
deoxyhemoglobin	2HHB		6.06	6.05	5.99	5.91

[a] f_d = experimental friction coefficient by diffusion. f_s = experimental friction coefficient by sedimentation. f_k = Kirkwood formula.

as determined in an extensive study of protein transport presented elsewhere.[15,16]

It is clear that the same picture emerges. The preaveraged approximation is quite good for translational diffusion, but it is poor for rotation (28% error) and the intrinsic viscosity (10% error). The BEST computations also agree quite well with experiment.[15,16]

A further comparison with a set of proteins that includes those previously computed by Teller[17] and co-workers (their Table 2) is shown in Table 6. In this case, only the translational friction coefficient is shown in order to compare with their computations.

We see again that for this entire set of proteins, the accurate BEST results and the preaveraged results differ by at most 1%. Preaveraging is quite accurate for irregular shapes. The comparison between BEST and the Kirkwood formula shows errors that range from 1.1% to 4.2%. Teller[17] et al. claimed that the discrepancy in the Kirkwood formula for translational friction for typical proteins was as large as 7%. However, our accurate computations show that the actual error of the Kirkwood formula is somewhat smaller. Our computations also show that the Kirkwood formula is not equivalent to preaveraging, as is well-known. The BEST values agree with experiment to within experimental error.[15]

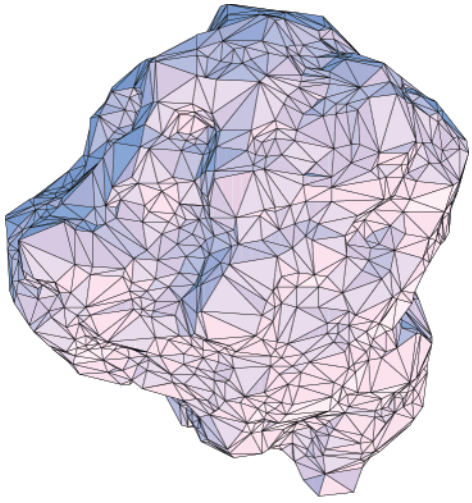


Figure 1. Triangulation of the hydrated surface of myoglobin.

IV. Conclusions

For the accurate computation of transport properties of rigid bodies, the hydrodynamic interaction should not be preaveraged except for the case of translational diffusion. For translational diffusion, the typical error in the preaveraging approximation is 1%, except for shapes with sharp corners, such as a tetrahedron, where the error rises moderately to about 3%. The errors are quite large for the rotational diffusion tensor, approaching 30%, and significant for the intrinsic viscosity, around 10%. The errors are shape dependent but take their typical magnitudes in the important case of globular proteins.

Acknowledgment. This research was supported through a grant from the National Institutes of Health, MBRS SCORE Program − Grant #S06 GM52588 to S.A. Helpful comments from Prof. J. M. Schurr are gratefully acknowledged.

References

(1) Happel, J.; Brenner, H. *Low Reynolds Number Hydrodynamics*; Prentice Hall: New York, 1965.

(2) Rotne, J.; Prager, S. Variational Treatment of Hydrodynamic Interaction in Polymers. *J. Chem. Phys.* **1969**, *50*, 4831−4837.

(3) Liu, B.; Dünweg, B. Translational diffusion of polymer chains with excluded volume and hydrodynamic interactions by Brownian dynamics simulation. *J. Chem. Phys.* **2003**, *118*, 8061−8072.

(4) Hubbard, B.; Douglas, J. F. Hydrodynamic friction of arbitrarily shaped Brownian particles. *Phys. Rev. E* **1993**, *47*, 2983−2986. Zhou, H.-X.; Szabo, A.; Douglas, J. F.; Hubbard, J. B. A Brownian dynamics algorithm for calculating the hydrodynamic friction and the electrostatic capacitance of an arbitrarily shaped object. *J. Chem Phys*. **1994**, *100*, 3821−3826.

(5) Douglas, J. F.; Garboczi, E. J. Intrinsic viscosity and the polarizability of particles having a wide range of shapes. *Adv. Chem. Phys.* **1995**, *91*, 85−133.

(6) Zhou, H. X. Calculation of translational friction and intrinsic viscosity. I. General formulation for arbitrarily shaped particles. *Biophys. J.* **1995**, *69*, 2286−2297. Zhou, H. X. A Unified Picture of Protein Hydration. *Biophys. Chem.* **2001**, *93*, 171−179.

(7) Garcia de la Torre, J.; Bloomfield, V. A. Hydrodynamic properties of complex, rigid biological macromolecules. Theory and Applications. *Quart. Rev. Biophys.* **1981**, *14*, 81−139.

(8) Aragon, S. A precise boundary element method for macromolecular transport properties. *J. Comput. Chem.* **2004**, *25*, 1191−1205.

(9) Youngren, G. K.; Acrivos, A. Stokes flow past a particle of arbitrary shape: a numerical method of solution. *J. Fluid Mech.* **1975**, *69*, 377−402.

(10) Allison, S. A.; Tran, V. T. Modeling the Electrophoresis of rigid polyions − Application to Lysozyme. *Biophys. J.* **1995**, *68*, 2261−2270.

(11) Allison. S. A.; Mazur, S. Modeling the free solution Electrophoretic mobility of short DNA fragments. *Biopolymers* **1998**, *46*, 359−373.

(12) Kim, S.; Karilla, S. J. *Microhydrodynamics*; Butterworth-Heinemann: New York, 1991.

(13) Simha, R. The Influence of Brownian Movement on the Viscosity of Solutions. *J. Phys. Chem.* **1940**, *44*, 25−34.

(14) We thank Prof. J. M. Schurr for the suggestion to consider this case.

(15) Aragon, S.; Hahn, D. K. Transport Properties of Proteins. I. Translational and Rotational Diffusion. *Biophys. J.* To be submitted.

(16) Hahn, D. K.; Aragon, S. Transport Properties of Proteins. II. Intrinsic Viscosity. *Biophys. J.* To be submitted.

(17) Teller, D. C.; Swanson, E.; De Haen, C. In *Methods in Enzymology*; Hirs, C. H. W., Timasheff, S. N., Eds.; Academic Press: New York, 1979: Vol. 61, pp 103−124.

(18) Richards, E. G. *An introduction to physical properties of large molecules in solution*; Cambridge University Press: London, 1980.

CT050158F

JCTC Journal of Chemical Theory and Computation

2D Entropy of Discrete Molecular Ensembles

J. Wang† and R. Brüschweiler*,‡

Carlson School of Chemistry and Biochemistry, Clark University, Worcester, Massachusetts, and Department of Chemistry and Biochemistry and National High Magnetic Field Laboratory, Florida State University, Tallahassee, Florida 32306

Received April 29, 2005

Abstract: A method is presented for the estimation of the conformational entropy of discrete macromolecular ensembles associated with multiple rotameric dihedral angle states. A covariance matrix is constructed of all mobile dihedral angles, which are represented as complex numbers on the unit circle, and subjected to a principal component analysis. The total entropy is decomposed into additive contributions from each eigenmode, for which a 2D entropy is computed after convolution of the projection coefficients of the conformer ensemble for that mode with a 2D Gaussian function. The method is tested for ensembles of linear polymer chains for which the exact conformational entropies are known. These include chains with up to 15 dihedral angles exhibiting two or three rotamers per dihedral angle. The performance of the method is tested for molecular ensembles that exhibit various forms of correlation effects, such as ensembles with mutually exclusive combinations of rotamers, ensembles with conformer populations biased toward compact conformers, ensembles with Gaussian distributed pairwise rotamer energies, and ensembles with electrostatic intramolecular interactions. For all these ensembles, the method generally provides good estimates for the exact conformational entropy. The method is applied to a protein molecular dynamics simulation to assess the effect of side-chain−backbone and side-chain−side-chain correlations on the conformational entropy.

1. Introduction

The thermodynamic stability of macromolecular states, such as ordered versus disordered states, is determined by their free energies, reflecting the balance between enthalpic and entropic contributions. Reliable estimates of entropy changes are, therefore, essential for the prediction and understanding of free energy changes.[1−9] For macromolecular systems, such as polymers and proteins, an important contribution is the conformational entropy. Unfortunately, the conformational entropy cannot be calculated analytically except for the simplest energy potentials. As an alternative, computer simulations are often used to sample relevant parts of the conformational space of macromolecules. Although such simulations produce discrete sets of conformers, the straightforward application of Boltzmann's equation $S = k \ln W$ to a computed trajectory with W snapshots generally bears little relevance with respect to the entropy. The ensemble of conformers first needs to be converted into probability distributions of relevant degrees of freedom before an entropy can be evaluated. In the quasiharmonic analysis method by Karplus and Kushick, the distribution of the various degrees of freedom of the discrete molecular ensemble, generated, for example, by molecular dynamics (MD) simulations, is approximated by a multivariate Gaussian distribution.[3,10] The quantity that enters the expression for the conformational entropy is the determinant of the covariance matrix of the coordinate fluctuations, which includes correlation effects between different degrees of freedom up to second order. Extensions of this approach have been developed that include

* Corresponding author phone: (850) 644−5173; fax: (850) 644−1366; e-mail: bruschweiler@magnet.fsu.edu.
† Clark University.
‡ Florida State University.

10.1021/ct050118b CCC: $33.50
Published on Web 11/03/2005

quantum-mechanical zero-point vibrational effects[12−17] and that address pure intramolecular reorientational entropic contributions.[18]

The quasiharmonic approximation does not always hold because the probability distribution of soft degrees of freedom, such as dihedral angles, is often significantly non-Gaussian as a result of anharmonic motions and the population of multiple rotameric states. Various methods that address these effects have been described.[19−23] In the method by Edholm and Berendsen, the conformational entropy is separately determined for each internal coordinate from the probability distribution of the ensemble along the coordinate by representing it as a histogram with a variable bin width.[20,21] A correction for correlation effects between internal coordinates is made by adding the difference of the quasiharmonic entropies in the presence and absence of correlations. More rigorous and computationally rather expensive alternative methods are Meirovitch's hypothetical scanning and local states methods that are capable of including correlation effects beyond second order (see ref 9 and references therein) and the method by Demchuk and coworkers that was applied to systems with one and two internal rotational degrees of freedom.[22,23]

In the present work, we describe a new method for estimating the conformational entropy of discrete molecular ensembles. It includes correlation effects up to second order in the complex representation $e^{i\varphi}$ of the molecule's internal torsion angles φ. A two-dimensional Gaussian distribution is assigned to each conformer along each eigenmode, and the conformational entropy is then determined as the sum of the entropy terms $-\int p(z) \log p(z)\, dz$ calculated along each mode. The method is first tested for a rotational isomeric state (RIS) model of polymer chains for which entropies can be determined analytically for reference. Different kinds of correlation effects are introduced to test the ability of the model to adequately reflect the entropy reduction associated with such effects. The model is finally applied to a MD trajectory of the protein ubiquitin.

2. Methods

We consider a linear polymer chain with N_a atoms connected by bonds of uniform length b and fixed bond angles. Each conformation (conformer) is fully specified by the $N_d = N_a - 3$ intervening dihedral angles φ_k, where $k = 1, ..., N_d$. The dihedral angles are represented as points on the unit circle in the complex plane, $z_k = e^{i\varphi_k}$, which circumvents the modulo 2π ambiguity of φ_k. Each conformer j is then specified by a vector $|d^{(j)}\rangle$:

$$|d^{(j)}\rangle = \{e^{i\varphi_1(j)}, e^{i\varphi_2(j)}, ..., e^{i\varphi_{N_d}(j)}\} \quad (1)$$

For an ensemble of N_c conformers, a complex covariance matrix $\mathbf{C}$ can be defined with elements

$$C_{kl} = \langle e^{i\varphi_k} e^{-i\varphi_l}\rangle - \langle e^{i\varphi_k}\rangle\langle e^{-i\varphi_l}\rangle,\ k, l = 1, ..., N_d \quad (2)$$

where the angular brackets indicate population-weighted averaging over the N_c conformers, for example, $\langle e^{i\varphi_k} e^{-i\varphi_l}\rangle = \sum_{j=1}^{N_c} p_j e^{i\varphi_k} e^{-i\varphi_l}$ where p_j is the population of conformer j with $\sum_j p_j = 1$, that is, for a conformational ensemble with a uniform distribution of populations $p_j = 1/N_c$. Using Euler's identity, the matrix elements of eq 2 can be expressed as

$$C_{kl} = \mathrm{cov}(\cos\varphi_k, \cos\varphi_l) + \mathrm{cov}(\sin\varphi_k, \sin\varphi_l) - \mathrm{icov}(\cos\varphi_k, \sin\varphi_l) + \mathrm{icov}(\sin\varphi_k, \cos\varphi_l) \quad (3)$$

where $\mathrm{cov}(f, g) = \langle f^* g\rangle - \langle f^*\rangle\langle g\rangle$.

A principal component analysis is then applied to matrix $\mathbf{C}$ by solving the eigenvalue problem $\mathbf{C}|m\rangle = \lambda_m|m\rangle$. The conformational entropy along each eigenmode $|m\rangle$ is calculated in the following way. First, each conformer $|d^{(j)}\rangle$ is projected along eigenmode $|m\rangle$, which yields the complex projection coefficients

$$c_{mj} = \langle m|d^{(j)}\rangle \quad (4)$$

The projection coefficients define the probability distribution of the conformational ensemble along mode m,

$$P_m(z)\, dz = \sum_{j=1}^{N_c} p_j \delta(z - c_{mj})\, dz \quad (5)$$

where $\delta(z - c_{mj}) = \delta[x - \mathrm{Re}(c_{mj})]\, \delta[y - \mathrm{Im}(c_{mj})]$ and where $\delta(x)$ is Dirac's delta function. Because of the finite number of conformers, $P_m(z)$ has a singular shape that is unsuitable for estimating entropies, and a smoothing procedure needs to be applied first.[24] The following procedure is used here: $P_m(z)$ is convoluted with a 2D Gaussian distribution with a standard deviation σ, $(2\pi\sigma^2)^{-1} \exp[-zz^*/(2\sigma^2)]$, which is normalized to ensure that the effective probability is constant. This yields the smoothed probability distribution

$$\tilde{P}_m(z)\, dz = \frac{1}{2\pi\sigma^2} \sum_{j=1}^{N_c} p_j \exp[-(z - c_{mj})(z^* - c_{mj}^*)/(2\sigma^2)]\, dz \quad (6)$$

σ is a smoothing parameter that needs to be calibrated in order to provide quantitative entropies as described below.

The entropy along mode m is then obtained by

$$S_m = -\int \tilde{P}_m(z) \ln \tilde{P}_m(z)\, dz \quad (7)$$

where the integral extends over the Gaussian plane. Here and in the following, Boltzmann's constant k_B is omitted; that is, all entropies are given in units of k_B unless noted otherwise. To correct for the net effect of the finite width σ on the entropy, a reference entropy S_{ref}, which is independent of m, is subtracted from S_m

$$S_{ref} = -\int \tilde{P}_{ref}(z) \ln \tilde{P}_{ref}(z)\, dz \quad (8)$$

where $P_{ref}(z)\, dz = (1/2\pi\sigma^2) \exp[-zz^*/(2\sigma^2)]\, dz$. Thus, the total entropy is obtained as

$$S_{2D} = \sum_{m=1}^{N_d} (S_m - S_{ref}) \quad (9)$$

Note that modes with eigenvalue $\lambda = 0$ do not yield a net contribution to S_{2D}. Because the entropy is computed from a distribution in the complex plane, that is, in two dimensions, it is termed S_{2D}. For the numerical evaluation of S_m

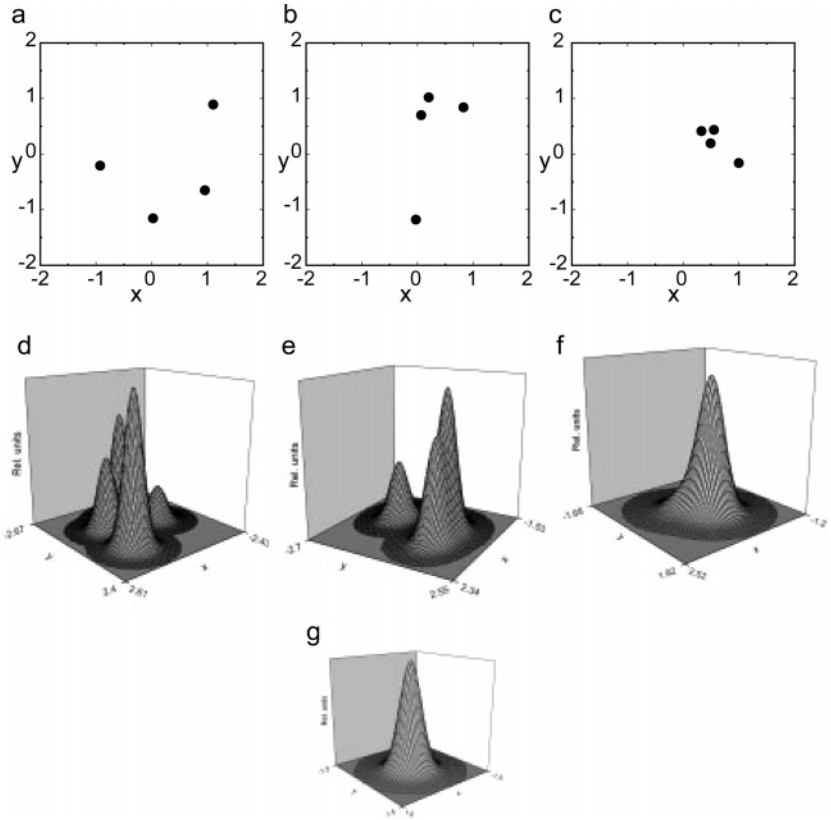


Figure 1. 2D probability distributions and projection coefficients of an ensemble of four conformers belonging to a linear chain molecule consisting of six atoms with three mobile dihedral angles. The dihedral angles of the four conformers are (φ_1, φ_2, φ_3) = (259.9°, 314.9°, 311.4°) for conformer 1, (19.9°, 74.9°, 311.4°) for conformer 2, (259.9°, 74.9°, 71.4°) for conformer 3, and (19.9°, 194.9°, 311.4°) for conformer 4, and their populations are $p_1 = 0.4$, $p_2 = 0.3$, $p_3 = 0.2$, and $p_4 = 0.1$, respectively. The covariance matrix **C** was calculated according to eq 2. Panels a−c show the complex projection coefficients c_{mj} (eq 4) for the three eigenvectors with nonzero eigenvalues $\lambda_1 = 1.243$, $\lambda_2 = 0.758$, and $\lambda_3 = 0.069$. The *x* and *y* axes correspond to the real and imaginary axes, respectively, of the complex plane. Panels d−f show the corresponding probability distributions calculated by the convolution of a 2D Gaussian function (panel g) with a standard deviation $\sigma = 0.3$ with the projection coefficients of panels a−c using eq 6.

and S_{ref}, in eqs 7 and 8, the integrals over the complex plane $z = x + iy$ are replaced by sums over a two-dimensional grid with boundaries at ±5 along *x* and *y* of each projection coefficient and a grid size of $\sigma/10$. Thus, for each mode, more than 10 000 grid points are evaluated. The 2D entropy can be compared with the analytical entropy

$$S_{\mathrm{a}} = -\sum_{j=1}^{N_{\mathrm{c}}} p_j \ln p_j \tag{10}$$

which, for uniform populations, $p_j = 1/N_{\mathrm{c}}$ is equivalent to Boltzmann's relationship $S_{\mathrm{a}} = \ln N_{\mathrm{c}}$.

3. Results

The entropy estimator described above is first tested for the RIS model of simple polymer chains for which the exact conformational entropy is known. In this model, the polymer is represented as a linear chain molecule consisting of N_{a} atoms with constant bond angles of 109.5° defined by consecutive atom triples. Each dihedral angle φ_k, which is defined by four consecutive atoms, occupies either $N_{\mathrm{r}} = 3$ or $N_{\mathrm{r}} = 2$ or rotameric states corresponding to a jump angle $\Delta\varphi$ of 120° and 180°, respectively. For each dihedral angle, the value of the first rotamer is either 0° (i.e., the four atoms defining the dihedral angle lie in the same plane forming a "cis" geometry) or it is chosen randomly between 0° and 360°. Excluded volume effects are considered by excluding any conformer for which one or more interatomic distances are shorter than the bond length *b*. For random values of the first rotamers, the total number of sterically allowed conformers may vary for different ensembles with the same number of dihedral angles.

An example of how discrete sets of projection coefficients are converted into continuous probability distributions is given in Figure 1. The projection coefficients are defined in eq 4, and the probability distributions used to evaluate the entropy are given in eqs 5−9. The figure shows the probability distributions for the three largest modes of a linear chain molecule consisting of $N_{\mathrm{a}} = 6$ atoms and $N_{\mathrm{d}} = 3$ dihedral angles with $N_{\mathrm{r}} = 3$. The generated ensemble consists of four conformers with conformer populations $p_1 = 0.4$, $p_2 = 0.3$, $p_3 = 0.2$, and $p_4 = 0.1$. Panels a−c display the projection coefficients for the three largest modes with $\lambda_1 = 1.243$, $\lambda_2 = 0.758$, and $\lambda_3 = 0.069$. Panels d−f show the corresponding probability distributions $\tilde{P}_m(z)$ after con-

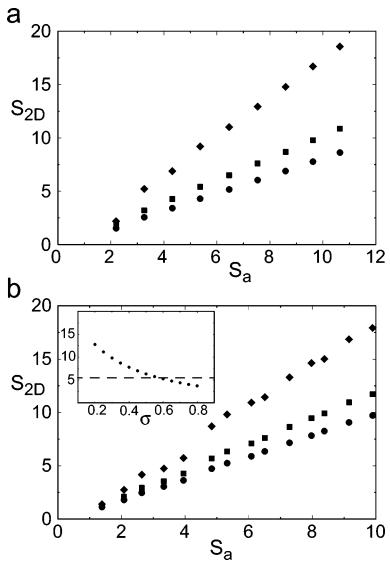


Figure 2. 2D entropy (eq 9) calculated for uniformly populated ensembles with $N_d = 2-10$ dihedral angles and $N_r = 3$ rotamers (panel a) and ensembles with $N_d = 2-14$ dihedral angles and $N_r = 2$ rotamers (panel b) vs the analytical entropy $S_a = \ln N_c$ (eq 10). The width σ was set to 0.3 (diamonds), 0.5 (squares), and 0.6 (circles). The insert in panel b shows the dependence of S_{2D} on σ for eight flexible dihedral angles (filled circles), and the horizontal line denotes S_a.

volution with the 2D Gaussian function with $\sigma = 0.3$ (see eq 6) of Panel g.

3.1. Uncorrelated Dihedral Angles. The effect of the Gaussian width σ on the 2D entropy S_{2D} is shown in Figure 2 for $N_r = 3$ and $N_r = 2$ as a function of the number of dihedral angles and in comparison to the analytical entropies S_a. σ is set to 0.3, 0.5, and 0.6. For a flip angle of $\Delta\varphi = 120°$ and $N_r = 3$ (panel a), the best agreement between S_{2D} and S_a is obtained for $\sigma = 0.5$, whereas for $N_r = 2$ (Panel b), the best agreement is obtained for $\sigma = 0.6$. The optimal value for σ shows a moderate dependence on the underlying rotameric jump model. The smaller the jump angle, the smaller is the optimal σ value because discrimination between the different rotameric states in the probability distribution $\tilde{P}_m(z)$ requires a narrower 2D Gaussian convolution function. The slight scatter in Figure 2 (as well as in Figure 3) is due to the random character of the first rotamer of each dihedral angle. A constant rotamer offset generally leads to smoother behavior.

3.2. Strongly Correlated Dihedral Angles. An essential criterion for the usefulness of an entropy estimator is that it faithfully takes into account the presence of correlations and anticorrelations between degrees of freedom. The behavior of S_{2D} was tested in this regard by generating ensembles with a reduced number of effective degrees of freedom by adding an increasing number of rotameric "mutual exclusivity constraints". Each such constraint precludes the simultaneous presence of two rotamers. For example, a constraint can impose that rotamer 1 of dihedral angle 5 is mutually exclusive with rotamer 3 of dihedral angle 7. Such constraints were randomly generated and successively applied to a 10-dihedral-angle ensemble with $N_r = 3$. The original ensemble consisting of 43 040 conformers that obey the excluded

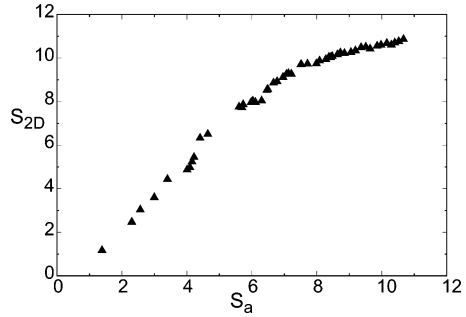


Figure 3. Reduction of entropies S_{2D} and S_a for an increasing number of mutually exclusive pairs of dihedral angles for a 10-dihedral-angle chain with $N_r = 3$. The total number of conformers is gradually reduced from 43 040 conformers in the absence of correlations (except for excluded volume effects) to 4 conformers upon introduction of an increasing number of pairwise correlations.

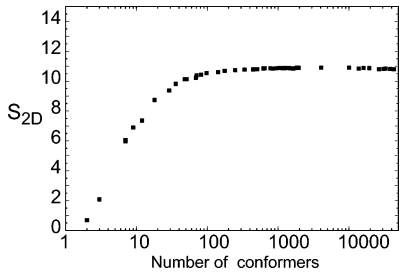


Figure 4. The effect of incomplete sampling on S_{2D} for a 10-dihedral-angle chain with $N_r = 3$. The initial ensemble of 43 161 allowed conformers was gradually reduced by excluding an increasing number of randomly selected conformers. S_{2D} with $\sigma = 0.5$ is plotted as a function of ensemble size N_c.

volume effect was thereby gradually reduced to an ensemble of only four conformers after adding up to 50 of these mutual exclusivity constraints. In Figure 3, for each ensemble, S_{2D} is plotted versus S_a. Overall, it shows a good equivalence between the two measures, although for small ensembles, S_{2D} tends to slightly overestimate the actual entropy. For ensembles constructed with other random sets of mutual exclusivity constraints, very similar relationships between S_a and S_{2D} are found.

3.3. Undersampling. In many practical (bio-)polymer applications, conformational space cannot be exhaustively searched, and a representative subset of conformational space is sampled instead. A good entropy estimator should allow extrapolation to the exact entropy from a relatively small subset of conformers. To test S_{2D} with respect to this property, a conformational ensemble is generated for the 10-dihedral-angle chain with $N_r = 3$. S_{2D} is then calculated for randomly chosen subsets of structures ranging between 2 and 43 161 conformers. A plot of S_{2D} versus the number of conformers N_c is given in Figure 4. It shows that S_{2D} converges rapidly toward the analytical entropy $S_a = 10.67$. A very good estimate of $S = 10.60$ is already obtained for 142 conformers, which accounts for less than 0.4% of all conformers. Since conformers are eliminated randomly, spurious correlations among dihedrals are mainly introduced in the limit of small numbers of conformers. This is in contrast to Figure 3 where significant correlations between

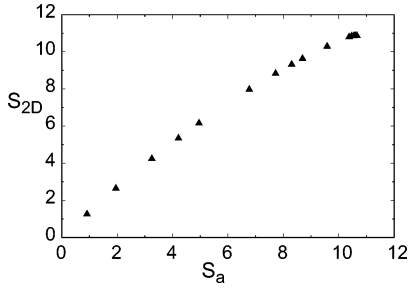


Figure 5. Correlation between S_{2D} and S_a for an ensemble of 10-dihedral-angle chains with $N_r = 3$ with conformer populations that are increasingly biased toward compact conformers as assessed by their radius of gyration: $p_k = (r_g - r_{g,min} + \Delta r_g)^{-1}$ (eq 11). Δr_g is varied between 10^{-6} and 1.0. The total number of allowed conformers is 43 082.

dihedral angles exist for any number of conformers. The convergence does not depend significantly on the value of σ for the range of interest here ($\sigma = 0.4-0.7$).

3.4. Soft Correlations: Radius of Gyration. In Figure 5, the behavior of S_{2D} is tested for ensembles whose conformer populations are biased toward conformers with a compact structure as reflected in a small radius of gyration. The conformer populations are given by

$$p_k = c(r_{g,k} - r_{g,min} + \Delta r_g)^{-1} \qquad (11)$$

where $r_{g,k}$ is the radius of gyration of conformer k computed as $r_{g,k}^2 = N_a^{-1} \sum_{j=1}^{N_a} \Delta r_{k,j}^2$, where $\Delta r_{k,j}$ is the distance of atom j to the center of mass of conformer k and c is a normalization constant. $r_{g,min}$ is the minimal radius of gyration of the ensemble, and Δr_g is an offset. The larger Δr_g, the more uniformly distributed are the probabilities, whereas for a small Δr_g, the most compact conformer dominates the ensemble.

In Figure 5, S_{2D} is compared with S_a for the 10-dihedral-angle conformational ensemble with $N_r = 3$ using the conformer probabilities of eq 11 with offset Δr_g varied between 10^{-6} (low entropy) and 1.0 (high entropy). The total number of conformers is 43 082, and the smallest and largest radii of gyration are 2.015 and 2.936, respectively. The good correlation between S_{2D} and S_a reflects the sensitive response of S_{2D} with respect to dihedral angle correlations underlying the global geometric property of compactness.

3.5. Soft Correlations: Gaussian Interaction Energies. A different method to introduce correlation effects between dihedral angles uses a pairwise energy potential function $E_{iu,jv}$, which denotes the energy between the uth rotamer of dihedral angle i and the vth rotamer of dihedral angle j. Using a Gaussian energy distribution

$$p(E_{iu,jv}) = \frac{1}{(2\pi\sigma_E^2)^{1/2}} e^{-(E_{iu,jv}-E_0)^2/(2\sigma_E^2)} \qquad (12)$$

where E_0 is an energy offset and σ_E is the standard deviation; the total energy of conformer k is given by $E_k = \sum_{i<j} \sum_{uv} E_{iu,jv}$, where the second sum goes over the rotamers occupied by conformer k; the relative population of a conformer k is given by $p_k = c \exp[-E_k/(k_B T)]$, where T is the absolute temperature. In Figure 6, S_{2D} is compared with S_a for the

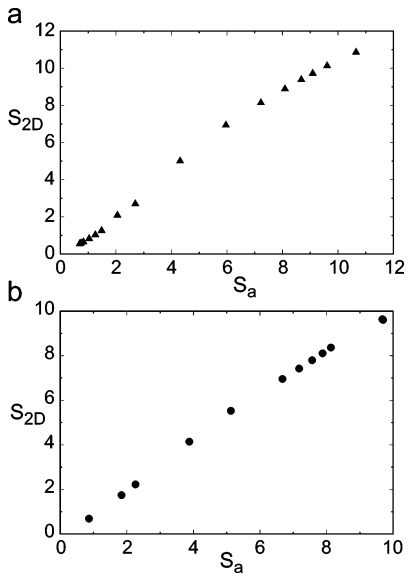


Figure 6. Comparison between S_{2D} and S_a for a Gaussian pairwise energy function with $k_B T$ varying between 0.1 and 1000. (Panel a) A 10-dihedral-angle chain with $N_r = 3$ and $\sigma = 0.5$. (Panel b) A 15-dihedral-angle chain with $N_r = 2$ and $\sigma = 0.6$.

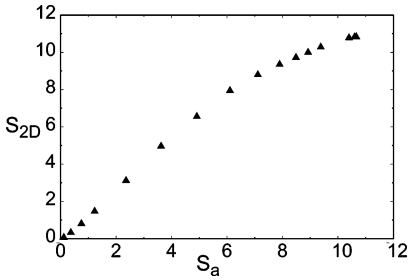


Figure 7. Comparison between S_{2D} and S_a for an ensemble of 10-dihedral-angle chains with $N_r = 3$ and $\sigma = 0.5$. Each atom carries a Coulomb charge of +1 or −1 with the total molecular charge being neutral. The entropies are calculated for Boltzmann distributions with different temperatures $k_B T$ ranging from 0.02 to 100.0.

10-dihedral-angle chain with $N_r = 3$ and $\sigma = 0.5$ (panel a) and the 15-dihedral-angle chain with and $N_r = 2$ and $\sigma = 0.6$ (panel b). For all calculations, E_0 is set to 10 and σ_E to 2, and $k_B T$ is varied from 0.1 (low entropy) to 1000 (high entropy). As can be seen from Figure 6, the correspondence between S_{2D} and S_a is good in both cases.

3.6. Soft Correlations: Electrostatic Energies. To test an alternative correlation mechanism, a distance-dependent interaction energy is used in the form of a Coulomb potential. For each atom, a charge of $q_i = 1$ or $q_i = -1$ was randomly assigned with the condition that the total charge of the molecule is zero. The corresponding Coulomb interaction takes the form $E = \sum_{i<j}(q_i q_j / r_{ij})$. The comparison between S_{2D} and S_a for an ensemble of 10-dihedral-angle chains with $N_r = 3$ and $\sigma = 0.5$ is depicted in Figure 7 with the temperature $k_B T$ varying between 0.02 (low entropy) and 100.0 (high entropy). Again, a good correspondence between the approximate and exact entropy measures can be seen, reflecting the sensitivity of S_{2D} to the correlation effects caused by the pairwise atomic Coulomb energy terms.

Table 1. S_{2D} of Ubiquitin in Units of J/mol/K

protein part	number of dihedrals	S_{2D} (correlated)	S_{2D} (uncorrelated)	ΔS_{2D}[a]
whole protein	312	483.01	561.08	78.07
backbone	151	129.73	134.65	4.91
side chains	161	372.63	426.44	53.81
backbone + side chains[b]	312	502.36[c]	561.09	58.72

[a] $\Delta S_{2D} = S_{2D}$ (uncorrelated) $- S_{2D}$ (correlated). [b] Arithmetic sums of backbone and side-chain entropies. [c] Includes all correlations, except correlations between backbone and side chains.

3.7. Application to Native Ubiquitin. The 2D entropy measure is applied to an ensemble of snapshots of the globular protein ubiquitin generated by molecular dynamics simulation. Despite the branched character of proteins, the entropy estimator is expected to deliver meaningful results with respect to the effect of dihedral angle correlations on the conformational entropy. The simulation was performed in a box with 2909 explicit water molecules at 300 K using CHARMM,[25] with details to be found elsewhere.[26,27] From a 5 ns trajectory, 1000 snapshots were extracted at a 5 ps time increment. The trajectory shows a stable behavior with a root-mean-square deviation of all heavy atoms (backbone and side chains) around 2 Å.[27] Since correlation times of some rotameric side-chain jump motions fall well into the nanosecond range, the total side-chain entropy is not fully converged for a simulation of this length. From each snapshot, all 151 mobile backbone φ, ψ dihedral angles were extracted, as well as the 161 mobile side-chain torsion angles, which amounts to a total of 312 dihedral angles. The 2D entropy is calculated as outlined in the Methods section using a standard deviation $\sigma = 0.5$ for the Gaussian convolution function. The results are summarized in Table 1. For the total conformational entropy S_{2D}, a value of 483.01 J/mol/K is obtained. The importance of correlation effects for S_{2D} is assessed by calculating S_{2D} after setting all off-diagonal elements to zero in covariance matrix **C**. This leads to an entropy increase of 78.07 J/mol/K. Backbone−backbone correlations contribute 4.91 J/mol/K, whereas side-chain−side-chain correlations contribute 53.81 J/ mol/K. The difference between 53.81 + 4.91 = 58.72 J/mol/K and 78.07 J/mol/K is −19.35 J/mol/ K, which reflects the entropy loss due to correlations between side-chain and backbone dihedral angles. As a consequence of dihedral angle correlations, the conformational entropy of the whole protein is reduced by 16.2%, for the backbone by 3.8% and for the side chains by 14.4%. Thus, motional side-chain−side-chain correlations are the dominant contributor to the conformational entropy loss in native ubiquitin.

4. Discussion

To gain useful insight into the thermodynamic properties of macromolecules from computer simulations (i) efficient sampling of conformational space and (ii) effective conversion of that information into thermodynamic quantities is required. The entropy measure, S_{2D}, introduced here represents a simple and robust estimator of the entropy associated with rotameric transitions of dihedral angles of an ensemble of conformers. Since dihedral angles are determined modulo 2π, there is an ambiguity in defining the dihedral angle average and its variance. This difficulty is avoided here by representing dihedral angles as complex numbers on the unit circle. Correlation effects between the dihedral angles are taken into account up to second order in terms of covariances and followed by a principal component analysis. Since this involves diagonalization of a complex $N_d \times N_d$ matrix, the method is efficient for systems with a small to moderately large number of dihedral angles. A continuous probability distribution is constructed for each eigenmode by convolution with a 2D Gaussian function with width σ. For an optimal choice of σ, information on the rotameric jump angles is required. This information can be obtained from the molecular force field or from the conformers themselves. The estimator is tested and calibrated on polymer chain models for which exact conformational entropies can be calculated for reference. The method provides good entropy estimates in the absence and presence of different types of correlation effects even when only a small fraction of all conformers is randomly sampled. S_{2D} focuses on the non-Gaussian dihedral angle distributions, reflecting primarily interconversion between different rotameric states. For these processes, the role of dihedral angle correlations is found in ubiquitin to be on the order of 16%. This contribution is dominated by motional correlations between side chains. Because of the finite width of σ and because in eq 9 the reference entropy S_{ref} is subtracted, dihedral angle variations that are significantly smaller than σ are not manifested in S_{2D}. σ can be viewed as a measure for the intrinsic structural uncertainty of a single conformer and thereby acts as a motional filter for the entropy evaluation: high-frequency vibrations and other small amplitude motions are not included in S_{2D} because their fluctuations are typically well-below the $\sigma = 0.5$ threshold. Such contributions can be evaluated using a normal-mode analysis[28−30] or quasiharmonic analysis applied to segments of MD or MC trajectories.[3−16]

Acknowledgment. This work was supported by NSF Grant MCB-0211512.

References

(1) Go, N.; Scheraga, H. A. *J. Chem. Phys.* **1969**, *51*, 4751−4767.

(2) Hagler, A. T.; Stern, P. S.; Sharon, R.; Becker, J. M.; Naider, F. *J. Am. Chem. Soc.* **1979**, *101*, 6842−6852.

(3) Karplus, M.; Kushick, J. N. *Macromolecules* **1981**, *14*, 325−332.

(4) Cheatham, T. E.; Srinivasan, J.; Case, D. A.; Kollman, P. A. *J. Biomol. Struct. Dyn.* **1998**, *16*, 265−280.

(5) Wrabl, J. O.; Shortle, D.; Woolf, T. B. *Proteins* **2000**, *38*, 123−133.

(6) Kuhn, B.; Kollman, P. A. *J. Med. Chem.* **2000**, *43*, 3786−3791.

(7) Schäfer, H.; Smith, L. J.; Mark, A. E.; van Gunsteren, W. F. *Proteins: Struct., Funct., Genet.* **2002**, *46*, 215−224.

(8) Gohlke, H.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 238−250.

(9) Cheluvaraja, S.; Meirovitch, H. *J. Chem. Phys.* **2005**, 122, 054903.

(10) Levy, R. M.; Karplus, M.; Kushick, J.; Perahia, D. *Macromolecules* **1984**, *17*, 1370−1374.

(11) Karplus, M.; Ichiye, T.; Pettitt, B. M. *Biophys. J.* **1987**, *52*, 1083−1085.

(12) Schlitter, J. *Chem. Phys. Lett.* **1993**, 215, 617−621.

(13) Brooks, B. R.; Janezic, D.; Karplus, M. *J. Comput. Chem.* **1995**, *16*, 1522−1542.

(14) Schäfer, H.; Mark, A. E.; van Gunsteren, W. F. *J. Chem. Phys.* **2000**, *113*, 7809−7817.

(15) Schäfer, H.; Daura, X.; Mark, A. E.; van Gunsteren, W. F. *Proteins* **2001**, *43*, 45−56.

(16) Andricioaei, I.; Karplus, M. *J. Chem. Phys.* **2001**, *115*, 6289−6292.

(17) Carlsson, J.; Aqvist, J. *J. Phys. Chem. B* **2005**, 109, 6448−6456.

(18) Prompers, J. J.; Brüschweiler, R. *J. Phys. Chem. B* **2000**, *104*, 11416−11424.

(19) Rojas, O. L.; Levy, R. M.; Szabo, A. *J. Chem. Phys.* **1986**, *85*, 1037−1043.

(20) Edholm, O.; Berendsen, H. J. C. *Mol. Phys.* **1984**, *51*, 1011−1028.

(21) Di Nola, A.; Berendsen, H. J. C.; Edholm, O. *Macromolecules* **1984**, *17*, 2044−2050.

(22) Hnizdo, V.; Fedorowicz, A.; Sing, H.; Demchuk, E. *J. Comput. Chem.* **2003**, *24*, 1172−1183.

(23) Darian, E.; Hnizdo, V.; Fedorowicz, A.; Sing, H.; Demchuk, E. *J. Comput. Chem.* **2005**, *26*, 651−660.

(24) In the quasiharmonic approximation,[3,10] for example, the δ peak distribution is "smoothed" by approximating it by a multivariate Gaussian function for which the entropy can be calculated analytically.

(25) Brooks, R. B.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.

(26) Lienin, S. F.; Bremi, T.; Brutscher, B.; Brüschweiler, R.; Ernst, R. R. *J. Am. Chem. Soc.* **1998**, *120*, 9870−9879.

(27) Prompers, J. J.; Scheurer, C.; Brüschweiler, R. *J. Mol. Biol.* **2001**, *305*, 1085−1097.

(28) Noguti, T.; Go, N. *J. Phys. Soc. Jpn.* **1983**, *52*, 3283−3288.

(29) Brooks, B. R.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 6571−6575.

(30) Levitt, M.; Sander, C.; Stern, P. S. *J. Mol. Biol.* **1985**, *181*, 423−447.

CT050118B

JCTC Journal of Chemical Theory and Computation

Stability of Carbon−Nitrogen Cages in 3-Fold Symmetry

Karleta D. Colvin,[†] Roshawnda Cottrell,[‡] and Douglas L. Strout*,[†]

Departments of Physical Sciences and Biological Sciences, Alabama State University, Montgomery, Alabama 36101

Received June 27, 2005

Abstract: Molecules consisting entirely of nitrogen have been studied extensively for their potential as high energy density materials (HEDM). One class of potential high-energy nitrogen molecules is the cage of three-coordinate nitrogen. Previous theoretical studies of cages N_x have shown that the most stable isomers are cylindrical molecules with 3-fold symmetry and triangular endcaps, but such molecules are not stable with respect to dissociation. In the current study, nitrogen cages are modified to include carbon atom substituents. Carbon atoms are studied for their potential to stabilize the nitrogen structures while maintaining significant levels of energy release from the molecules. Theoretical calculations are carried out on a sequence of high-energy cages with carbon and nitrogen. Density functional theory (B3LYP), perturbation theory (MP2 and MP4), and coupled-cluster theory (CCSD(T)) are used in conjunction with the correlation-consistent basis sets of Dunning. Stability trends as a function of molecule size are calculated and discussed.

Introduction

Nitrogen molecules have been the subject 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 heteroatoms and polyazido isomers[18] of compounds 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[19] 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[20] 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[21] 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[22,23] of intermediate-sized molecules N_{14}, N_{16}, and N_{18} also showed that the cage

* Corresponding author phone: (334)229-4718; e-mail: dstrout@alasu.edu.

[†] Department of Physical Sciences.

[‡] Department of Biological Sciences.

10.1021/ct050163j CCC: $33.50
Published on Web 10/29/2005

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] 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[24,25] or phosphorus substitution.[26] The phosphorus study predicted the stability of a molecule of N_6P_6, but phosphorus is a high-mass atom that does not contribute appreciably to energy release. These atoms dilute the energy-per-unit-mass properties of the molecule. Therefore, in designing a viable HEDM, it is not only necessary to have a stable molecule but also desirable to minimize the number and mass of heteroatoms and thereby maximize energy production from the HEDM. In the current study, several molecules are studied whose structures are based on the most stable N_{12} and N_{18} but with carbon atoms (much lighter than phosphorus) substituted into the cage network. The stability of carbon−nitrogen cages is determined by theoretical calculations of the energies of various dissociation pathways of each molecule.

Computational Details

Geometries are optimized with density functional theory[27,28] (B3LYP) and second-order perturbation theory[29] (MP2). Single energy points are calculated with fourth-order perturbation theory[29] (MP4(SDQ)) and coupled-cluster theory[30] (CCSD(T)). Multireference effects are calculated by complete active space (CASSCF(4,4)) calculations with MP2 energies included. Molecules are optimized in the singlet state, and dissociation intermediates are optimized in the triplet state, which is the ground state for all dissociations in this study. The basis sets are the double-ζ (cc-pVDZ), augmented double-ζ (aug-cc-pVDZ), and triple-ζ (cc-pVTZ) sets of Dunning.[31] Vibrational frequencies have been calculated at the MP2/cc-pVDZ level of theory for $N_6C_6H_6$ and for all its dissociation intermediates. For the larger intact molecules, B3LYP/cc-pVDZ frequencies have been calculated. The Gaussian03 computational chemistry software,[32] and its Windows-based counterpart Gaussian03W, have been used for all calculations in this study.

Results and Discussion

The first molecule under consideration in this study is a variation on the most stable N_{12} cage. The molecule has two triangular endcaps that have been replaced by carbon atoms. Including the hydrogens that are added for the fourth bond of carbon, this molecule has the formula $N_6C_6H_6$ and is shown in Figure 1. The molecule has D_{3d} point group symmetry and three symmetry-independent bonds, and the intermediates for breaking each bond are shown in Figures 2−4. The dissociation energies for bond-breaking processes of $N_6C_6H_6$ are shown in Table 1. (The molecule and all of its one-bond-breaking intermediates are been verified as local minima, and the effects of zero-point energy and free energy are shown in Table 1.) MP2 and B3LYP energies do not agree, with B3LYP giving lower bond dissociation energies. The most easily broken bond is the nitrogen−nitrogen (NN) bond, but even this bond has a dissociation energy of more than 30 kcal/mol at the CCSD(T)/cc-pVDZ level of theory, which is the most reliable method in this study. The MP4/cc-pVDZ results agree closely with CCSD(T). Basis set effects from diffuse functions (aug-cc-pVDZ) or higher angular momentum functions (cc-pVTZ) tend to increase the bond dissociation energies. Multireference effects on the dissociation energies have been calculated by CASSCF(4,4) calculations with MP2 corrections, resulting in increases in the dissociation energies by 5−8 kcal/mol. Since all of the bonds in the $N_6C_6H_6$ have high dissociation energies, this molecule is probably a good candidate for a practical HEDM. However, this molecule is only 52% nitrogen by mass, and since nitrogen is the source of the energy release, it would be desirable to increase the percentage of nitrogen in the molecule if possible while maintaining stability.

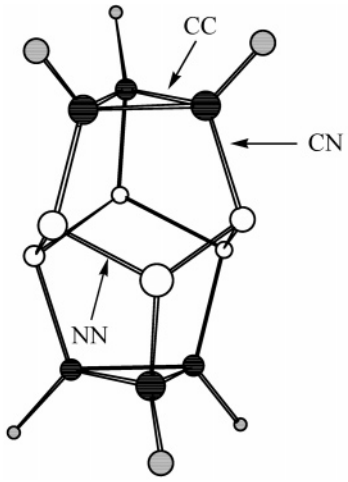


Figure 1. $N_6C_6H_6$ molecule (D_{3d} point group symmetry). Symmetry-independent bonds are labeled. Nitrogen is shown in white, carbon in black, and hydrogen in gray.

Table 1. Bond-Breaking Energies for $N_6C_6H_6$ Molecule[a]

		bonds (see Figure 1)		
energy	geometry	CC	CN	NN
B3LYP/cc-pVDZ	B3LYP/cc-pVDZ	+68.4	+85.8	+21.3
MP2/cc-pVDZ	MP2/cc-pVDZ	+78.7	+105.5	+42.5
MP2(+ZPE)/cc-pVDZ	MP2/cc-pVDZ	+76.6	+102.5	+41.5
MP2(+free energy)/cc-pVDZ	MP2/cc-pVDZ	+74.6	+100.3	+39.7
CAS(4,4)MP2/cc-pVDZ	MP2/cc-pVDZ	+83.3	+111.8	+49.9
MP2/aug-cc-pVDZ	MP2/aug-cc-pVDZ	+78.5	+106.5	+44.9
MP2/cc-pVTZ	MP2/cc-pVTZ	+81.7	+109.2	+45.2
MP4/cc-pVDZ	MP2/cc-pVDZ	+71.3	+96.1	+31.8
MP4/aug-cc-pVDZ	MP2/aug-cc-pVDZ	+70.6	+96.3	+33.3
CCSD(T)/cc-pVDZ	MP2/cc-pVDZ	+71.4	+93.0	+31.4

[a] Energies in kcal/mol.

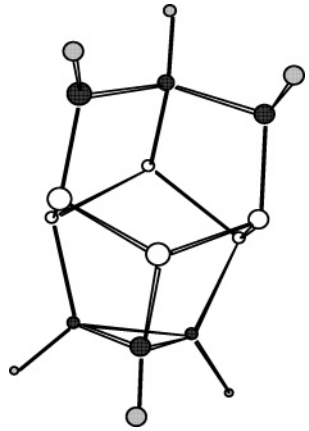


Figure 2. $N_6C_6H_6$ molecule, with a C−C bond broken. Nitrogen is shown in white, carbon in black, and hydrogen in gray.

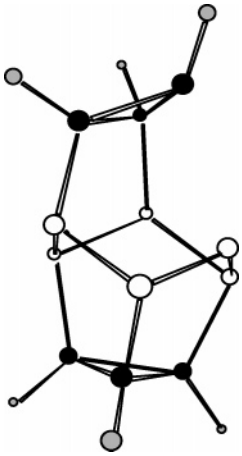


Figure 3. $N_6C_6H_6$ molecule, with a C−N bond broken. Nitrogen is shown in white, carbon in black, and hydrogen in gray.

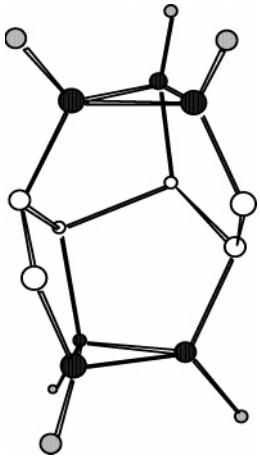


Figure 4. $N_6C_6H_6$ molecule, with a N−N bond broken. Nitrogen is shown in white, carbon in black, and hydrogen in gray.

It is possible to design a molecule with the same carbon end-caps with two six-membered rings of nitrogen instead of only one. This molecule has a formula $N_{12}C_6H_6$ and is shown in Figure 5. This molecule is 68% nitrogen by mass and more energetic than $N_6C_6H_6$ as shown in Table 2, but is

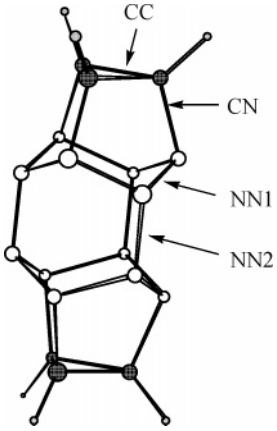


Figure 5. $N_{12}C_6H_6$ molecule (D_{3h} point group symmetry). Symmetry-independent bonds are labeled. Nitrogen is shown in white, carbon in black, and hydrogen in gray.

Table 2. Free Energies of Reaction for Molecules in This Study, Calculated by the B3LYP/cc-pVDZ Method

molecule	reaction	kcal/mol	kcal/g
$N_6C_6H_6$	$N_6C_6H_6 \rightarrow 3N_2 + C_6H_6$	−271.0	−1.7
$N_{12}C_6H_6$	$N_{12}C_6H_6 \rightarrow 6N_2 + C_6H_6$	−561.8	−2.3
$N_{12}C_9H_6$	$N_{12}C_9H_6 \rightarrow 6N_2 + (1/2)C_6H_6 + (1/4)C_{24}H_{12}$	−555.1	−2.0
$N_{18}C_{12}H_6$	$N_{18}C_{12}H_6 \rightarrow 9N_2 + (1/2)C_{24}H_{12}$	−840.8	−2.1
N_6P_6	$N_6P_6 \rightarrow 3N_2 + (3/2)P_4$ (ref 26)	−230.8	−0.9

Table 3. Bond-Breaking Energies for $N_{12}C_6H_6$ Molecule[b]

		bonds (see Figure 5)			
energy	geometry	CC	CN	NN1	NN2
B3LYP/cc-pVDZ	B3LYP/cc-pVDZ	+70.6	+76.8	*a*	*a*
MP2/cc-pVDZ	MP2/cc-pVDZ	+80.8	+104.0	+30.5	+86.2
MP4/cc-pVDZ	MP2/cc-pVDZ	+73.4	+95.6	+23.1	+75.3
MP2/aug-cc-pVDZ	MP2/aug-cc-pVDZ	+80.6	+105.1	+33.7	+88.9

[a] Geometry optimization was unsuccessful. [b] Energies in kcal/mol.

it stable with respect to dissociation? Bond-breaking energies are shown in Table 3 for the four symmetry-independent bonds (in D_{3h} symmetry). As with $N_6C_6H_6$, the weakest bond is the nitrogen−nitrogen bond (NN1) within a ring of nitrogen (as opposed to NN2, which connects the two rings of nitrogen). The bond-breaking energy is much lower than in the smaller molecule, 12 kcal/mol lower at the MP2/cc-pVDZ level of theory, and 9 kcal/mol at the MP4/cc-pVDZ level of theory. At the highest level of theory in this study, the molecule has less than 30 kcal/mol resistance to dissociation and is likely only a marginal candidate for HEDM. This is an effect similar to what was shown[25] for a series of carbon−oxygen-capped molecules with stacked six-membered rings of nitrogen. It seems that ring-stacking nitrogen upon nitrogen leads to weakening of N−N single bonds for large nitrogen cage structures.

Would separating the two nitrogen rings result in a stability enhancement for the molecule? Figure 6 shows a molecule with another triangle of carbon between the two rings of nitrogen. This molecule has the formula $N_{12}C_9H_6$, which is 60% nitrogen by mass. The molecule has D_{3h} point group symmetry and five symmetry-independent bonds. The dis-

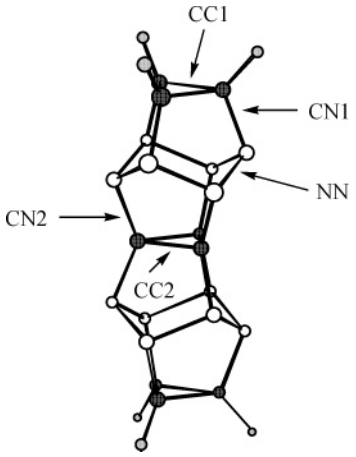


Figure 6. $N_{12}C_9H_6$ molecule (D_{3h} point group symmetry). Symmetry-independent bonds are labeled. Nitrogen is shown in white, carbon in black, and hydrogen in gray.

Table 4. Bond-Breaking Energies for $N_{12}C_9H_6$ Molecule[b]

		bonds (see Figure 6)				
energy	geometry	CC1	CN1	NN	CN2	CC2
B3LYP/cc-pVDZ	B3LYP/cc-pVDZ	*a*	+89.4	+15.0	+71.1	+70.1
MP2/cc-pVDZ	MP2/cc-pVDZ	+78.8	+108.8	+36.7	+89.1	+86.0
MP4/cc-pVDZ	MP2/cc-pVDZ	+71.1	+98.5	+25.3	+79.6	+76.2

[a] Geometry optimization was unsuccessful. [b] Energies in kcal/mol.

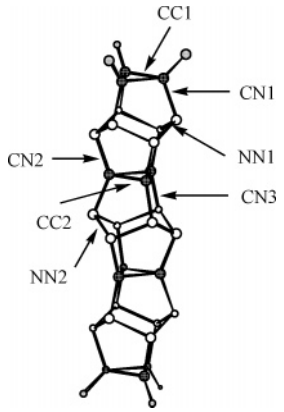


Figure 7. $N_{18}C_{12}H_6$ molecule (D_{3d} point group symmetry). Symmetry-independent bonds are labeled. Nitrogen is shown in white, carbon in black, and hydrogen in gray.

sociation energies for each bond are shown in Table 4. For the weakest bond in both molecules ($N_{12}C_6H_6$ and $N_{12}C_9H_6$), which is a nitrogen−nitrogen bond, the additional ring of carbon results in 6 kcal/mol of additional stability at the MP2/cc-pVDZ level of theory (+36.7 versus +30.5 kcal/mol dissociation energy). At the MP4/cc-pVDZ level of theory, the stability advantage of the additional ring of carbon diminishes to about 2 kcal/mol (+25.3 versus +23.1 kcal/mol).

The structure of $N_{12}C_9H_6$ can be extended with another ring of six nitrogens and a ring of three carbons to form a larger molecule with formula $N_{18}C_{12}H_6$. This molecule is shown in Figure 7 and consists of 63% nitrogen by mass. The molecule has seven symmetry-independent bonds between heavy atoms, and the dissociation energies for the bonds are shown in Table 5. The results indicate that the weakest bond in the molecule is a nitrogen−nitrogen bond in the central ring of nitrogen atoms. Comparing MP2/cc-pVDZ results with the smaller $N_{12}C_9H_6$ reveals that the molecule with three rings of nitrogen is less stable than the molecule with two. $N_{18}C_{12}H_6$ can break an N−N bond more easily (by about 6 kcal/mol) than $N_{12}C_9H_6$. The $N_{18}C_{12}H_6$ molecule is therefore unlikely to be a stable HEDM.

Table 5. Bond-Breaking Energies for $N_{18}C_{12}H_6$ Molecules[b]

bonds (see Figure 7)	B3LYP/cc-pVDZ	MP2/cc-pVDZ
CC1	*a*	+78.8
CN1	+89.6	+109.0
NN1	+15.1	+36.7
CN2	+70.7	+88.9
CC2	+68.9	+85.6
CN3	+74.5	+92.2
NN2	+8.0	+30.1

[a] Geometry optimization was unsuccessful. [b] Energies in kcal/mol.

Conclusion

Carbon is a viable heteroatom substituent in stabilizing N_{12} to form the stable $N_6C_6H_6$. However, lengthening schemes designed to extend the stabilizing features of $N_6C_6H_6$ to larger, nitrogen-richer molecules result in molecules that are less stable than $N_6C_6H_6$. These larger molecules are therefore less likely to serve as a practical high energy density material (HEDM). As a substitute for nitrogen, the lighter carbon atoms have a less drastic effect on energy output than the heavier, previously studied phosphorus atom substituents. The $N_6C_6H_6$ molecule is stable enough to serve as an HEDM, and it should have energy release properties much more favorable than N_6P_6.

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, IL. K.D.C. and R.C. are undergraduate scholars supported by the Minority Access to Research Careers (MARC) program administered by the National Institute of General Medical Sciences (NIH/NIGMS 5T34GM08167-20). D.L.S. is supported by MARC as a faculty research mentor. This work was also supported by the National Institutes of Health (NIH/NCMHD 1P20MD000547-01). The taxpayers of the state of Alabama in particular and the United States in general are gratefully acknowledged.

Supporting Information Available: MP2/cc-pVDZ optimized geometries for molecules and intermediates (coordinates in Å). 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., III *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) Li, Q. S.; Liu, Y. D. *Chem. Phys. Lett.* **2002**, *353*, 204. Li, Q. S.; Qu, H.; Zhu, H. S. *Chin. Sci. Bull.* **1996**, *41*, 1184.

(7) Li, Q. S.; Zhao, J. F. *J. Phys. Chem. A* **2002**, *106*, 5367. 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) Vij, A.; Pavlovich, J. G.; Wilson, W. W.; Vij, V.; Christe, K. O. *Angew. Chem., Int. Ed.* **2002**, *41*, 3051. 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. *Nature 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) Glukhovtsev, M. N.; Jiao, H.; Schleyer, P.v. R. *Inorg. Chem.* **1996**, *35*, 7124.

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

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

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

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

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

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

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

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

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

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

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

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

(32) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03, revision B.01*; Gaussian, Inc.: Wallingford, CT, 2003.

CT050163J

JCTC Journal of Chemical Theory and Computation

Quantifying Aromaticity at the Molecular and Supramolecular Limits: Comparing Homonuclear, Heteronuclear, and H-Bonded Systems

Abdul Rehaman, Ayan Datta, Sairam S. Mallajosyula, and Swapan K. Pati*

Theoretical Sciences Unit and Chemistry and Physics of Materials Unit, Jawaharlal Nehru Centre for Advanced Scientific Research, Jakkur Campus, Bangalore 560 064, India

Received June 22, 2005

Abstract: The aromatic/antiaromatic characteristics of B−N and P−N analogues of benzene and cyclobutadiene have been studied using quantum chemical methods. We use established parameters such as nucleus-independent chemical shifts, charge density at the ring critical point, and stabilization energies to quantify the nature of interactions in these molecular systems. $B_3N_3H_6$ and $N_3P_3F_6$ resemble benzene in being aromatic, albeit to a lesser extent, while $B_2N_2H_4$ and $N_2P_2F_4$ are found to be aromatic, opposite to that for cyclobutadiene. A $\sigma-\pi$ separation analysis has been performed to critically examine the contributions from the π electrons compared to that from the σ backbone. The structural aspects in the weak interaction limits such as the H-bonded cyclic trimers of HX (X = F, Cl, and Br) have also been investigated. Even in such weak interaction limits, these cyclic systems are found to be substantially stable. These H-bonded systems exhibit nonlocal polarizations across the full-perimeter of the ring that lead to aromaticity. We propose the term "*H-bonded aromaticity*" for such closed-loop weakly delocalized systems. This new formalism of aromaticity has the potential to explain structures and properties in supramolecular systems.

I. Introduction

Aromaticity is a well-known and useful concept in organic chemistry. Though the initial interpretation of aromaticity was based only on one-electron theories such as the Hückel model, modern quantum chemical calculations have now established this concept on a firm footing.[1] The rapid synthesis and characterization of new molecules exhibiting aromaticity/antiaromaticity have further fueled the interest in these systems.[2,3] From the conventional rules of aromaticity/antiaromaticity in organic molecules, the concept has been recently introduced to all-metal clusters with the proposal of d-orbital aromaticity and σ aromaticity.[4−6] The past decade has also witnessed a renewed interest in concepts such as Möbius aromaticity with the synthesis of molecular Möbius systems.[7] Also in the same context, three-dimensional aromaticity in molecular complexes has led to the stabilization of many otherwise unstable molecular systems[8] and organometallic sandwich complexes, for example, dinuclear Zn complexes ($Cp^*_2-Zn_2$) (Cp = cyclopentadiene).[9]

Central to the concept of aromaticity and antiaromaticity is the simple yet widely successful Hückel rule that predicts $(4n + 2)\pi$ planar electronic systems to be aromatic and stable, while $4n\pi$ electronic systems are antiaromatic and unstable. The Hückel rule is very successful in representative organic molecules such as benzene and cyclobutadiene. It is also quite applicable to heterocyclic organic molecules.[10] However, the application of the rule cannot be extended to the realms of inorganic molecules. The aromaticity in inorganic molecules such as borazine and phosphazene has been a long-debated issue. For example, though $B_3N_3H_6$ and $N_3P_3F_6$ have a resemblance to benzene, both in structure and reactivity, these molecules differ substantially from benzene, and such differences have been widely reported in the literature.[11] However, for the four-membered ring systems

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

10.1021/ct0501598 CCC: $33.50
Published on Web 11/09/2005

such as $B_2N_2H_4$ and $N_2P_2F_4$, synthesis has been quite difficult, and only a few four-membered ring systems with sterically hindered ligands have been realized so far.[12] Thus, in the $4n\pi$ manifold of these charge transfer (CT) complexes, the structure−property relationship is yet to be fully understood.

For the present work, we consider various four- and six-membered rings of homonuclear and heteronuclear systems. These rings are either stabilized through a complete delocalization of the π electrons (as for homonuclear systems) or through a partial or complete charge transfer of π electrons due to electronegativity differences between the atoms (as for heteronuclear systems). We also consider a class of cyclic systems where the stabilization is due to weak hydrogen-bonding interactions. H-bonded interactions are known to follow directionality,[13,14] and on the basis of graph theory analyses,[15,16] it has been suggested that cyclic polygonal closed-loop structures are stable geometries for H-bonded molecules. Similar conclusions are also derived from modern quantum chemical calculations.[17] The nature of σ-electron delocalizations and, thus, aromaticity has been critically examined for such molecules. We, thus, compare and contrast the aromaticity or the lack of it in complexes ranging from purely covalent, to ionic, to partially ionic, to weakly interacting systems.

In the next section, we perform ab initio calculations on the homoatomic (C_6H_6 and C_4H_4) and heteroatomic ($B_3N_3H_6$, $N_3P_3F_6$, $B_2N_2H_4$, and $N_2P_2F_4$) systems. Following the calculations, we critically examine the aromaticity/antiaromaticity characteristics for these systems along with the weakly interacting H-bonded systems. We then critically examine the role of the delocalized π electrons and the σ framework in rationalizing the stability for the molecular structures. Finally, we conclude the paper with a summary of the results.

II. Computational Details and Results

All the geometries for the molecular systems considered in this work have been fully optimized at the density functional theory (DFT) level using the Becke, Lee, Yang, and Parr three-parameter correlation functional (B3LYP) at the 6-311G++(d,p) basis set level.[18] All the calculations have been performed using the Gaussian 03 set of programs.[19] Additional calculations at the MP2 level have been performed to further verify the structures for these molecules. Also, frequency calculations are performed to confirm the ground-state structures (see Supporting Information).

In Figure 1, the optimized structures for all the π-delocalized systems are shown. It is seen that, for the homoatomic systems, C_6H_6 and C_4H_4 (Figure 1a and b), the bond-length alterations (BLAs; defined as the average difference between the bond lengths of two consecutive bonds) are 0.00 and 0.24 Å, representing aromatic and antiaromatic features, respectively. The six-membered heteroatomic clusters, $B_3N_3H_6$ and $N_3P_3F_6$ (Figure 1c and e), have a 0 BLA high-symmetric hexagonal structure. For the four-membered rings, $B_2N_2H_4$ and $N_2P_2F_4$ (Figure 1d and f), the structures are rhombohedral with equal bond lengths and unequal diagonal lengths. For $B_2N_2H_4$, the shorter and longer diagonals are 1.90 and 2.15 Å, respectively, and for $N_2P_2F_4$, they are 2.15 and 2.47 Å, respectively. Thus, for the homoatomic C_4H_4, John−Teller

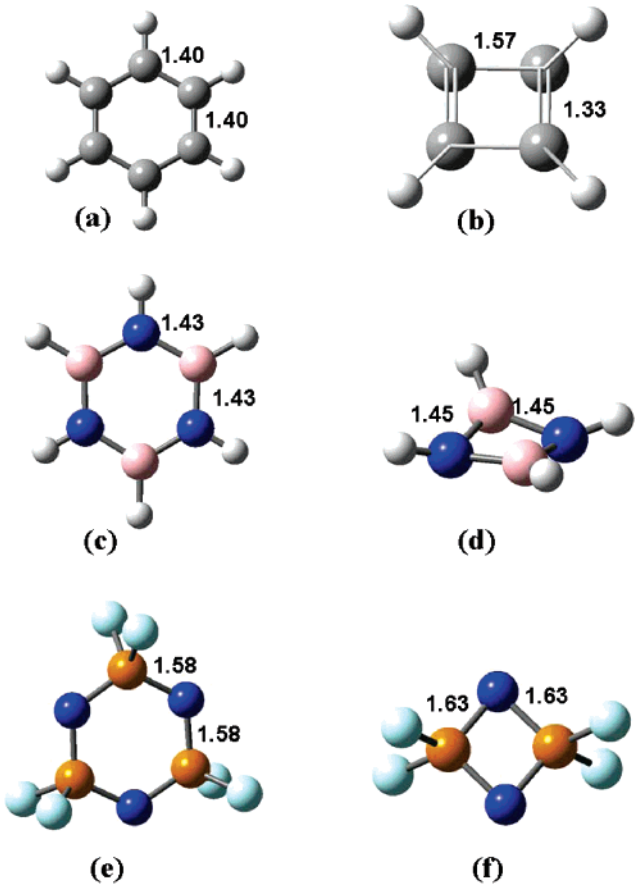


Figure 1. Ground-state optimized geometries of (a) C_6H_6, (b) C_4H_4, (c) $B_3N_3H_6$, (d) $B_2N_2H_4$, (e) $N_3P_3F_6$, and (f) $N_2P_2F_4$. Bond lengths in Å shown for each structure.

(JT) distortion leads to a rectangular geometry from a square geometry, while for the heteroatomic four-membered ring systems, such distortions lead to a rhombohedral geometry.

In Figure 2, we show the highest occupied molecular orbitals (HOMO) for the above-mentioned molecular systems. For the homoatomic molecular systems such as C_6H_6 and C_4H_4 (Figure 2a and b), the nodal plane passes through the bonds and the MO is delocalized over each atom. But, for $B_3N_3H_6$ and $N_3P_3F_6$ (Figure 2c and e), the MOs are indicative of electronegativity differences. For the four-membered ring systems, $B_2N_2H_4$ and $N_2P_2F_4$ (Figure 2d and f), very large contributions are observed for the N atoms and negligible contributions from the electropositive atoms are observed, suggesting CT from the N orbital to the vacant orbitals (p_z orbital of B and d_{xz} and d_{yz} orbitals of P). As expected on the basis of symmetry and relative electronegativities, for $B_2N_2H_4$ (Figure 2d), the node passes through the less electronegative B atoms. In sharp contrast, the node passess through the longer C−C bonds for C_4H_4 (Figure 2b).

The case of the four-membered B−N compound, $B_2N_2H_4$, requires a special mention. The ground-state structure corresponds to a puckering of 17.3° from planarity. However, the bond lengths are all equivalent, suggesting that the lone pair of electrons on the N atom are localized and are not transferred to the nearby B atom, and a resonance form similar to that for C_4H_4 (two alternate short and long bonds) is not realized. In fact, the planar structure for $B_2N_2H_4$ is

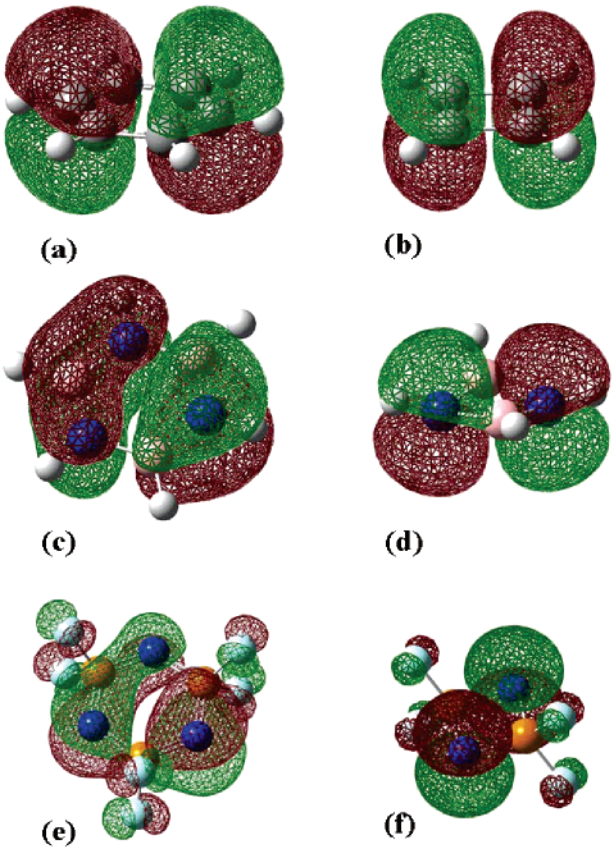


Figure 2. Highest occupied molecular orbitals (HOMOs) of (a) C_6H_6, (b) C_4H_4, (c) $B_3N_3H_6$, (d) $B_2N_2H_4$, (e) $N_3P_3F_6$, and (f) $N_2P_2F_4$.

1.00 kcal/mol higher in energy compared to the puckered structure and has one imaginary frequency corresponding to the out-of-plane bending mode of the atoms. However, a difference of 1 kcal/mol in energies between these two structures is comparable to the thermal energy at room temperature (0.6 kcal/mol). Thus, there is a possibility for such four-membered rings to exist in two different polymorphs: the planar and the puckered forms. We performed a search for such polymorphs in the Cambridge Crystallographic Database[20] (CCSD) for the four-membered B_2N_2 ring systems and had 47 hits. Of them, two compounds show polymorphism in the ring structure. For example, the crystal of 1,3-di-*tert*-butyl-2,4-bis(pentafluorophenyl)-1,3,2,4- diazadiboretidine (CCSD code: BFPDZB) crystallizing in an *I*2/*c* point group has a planar B_2N_2 unit,[21] while the tetrakis-(*tert*-butyl)-1,3,2,4-diazadiboretidine crystal (CCSD code: CETTAW) with a point group of *Pc* maintains a puckered B_2N_2 unit with a puckering angle of 18°.[22] Note that our computed structure also has a similar puckering angle. Thus, the existence of the two crystal polymorphs in the B_2N_2 unit strongly support our calculations.

III. Aromaticity Criteria

For a quantitative measurement of aromaticity/antiaromaticity in these systems, we have calculated the nucleus-independent chemical shifts (NICS) at the center of each ring structure. Compounds with exalted diamagnetic susceptibility are aromatic, while those showing paramagnetic susceptibilities are antiaromatic.[23] Also, another parallel method to characterize aromaticity/antiaromaticity is to calculate the charge density (ρ_{RCP}) and its Laplacian ($\nabla^2\rho_{RCP}$) at the ring critical point. In general, molecules with similar architecture share similar topological features and, thus, serve as a tool for understanding structural aspects in molecules. There have been intense efforts to relate these topological aspects with aromaticity/antiaromaticity criteria recently.[24]

Table 1. Magnitudes of Bond Length Alteration (BLA) in Å, Stabilization Energies in kcal/mol (ΔE), Nucleus-Independent Chemical Shift (NICS) in ppm, Charge Density at the Ring Critical Point (ρ_{RCP}) in e/Å^3 Units, and Laplacian of the Charge Density ($\nabla^2\rho_{RCP}$) in e/Å^5 Units for the Systems Considered in the Present Study

systems	BLA	ΔE	NICS	ρ_{RCP}	$\nabla^2\rho_{RCP}$
C_6H_6	0.00	−219.36	−8.072	0.022	0.161
C_4H_4	0.24	−52.29	35.763	0.102	0.458
$B_3N_3H_6$	0.00	−142.16	−1.595	0.020	0.119
$B_2N_2H_4$	0.00	−49.35	−2.921	0.099	0.369
$N_3P_3F_6$	0.00	−238.51	−6.912	0.020	0.095
$N_2P_2F_4$	0.00	−105.61	−10.874	0.104	0.210
$(HF)_3$	0.90	−10.94	−2.94	0.008	0.046
$(HCl)_3$	1.29	−4.23	−1.98	0.002	0.006
$(HBr)_3$	1.35	−1.67	−1.89	0.002	0.006

In Table 1, we tabulate the magnitudes of the BLA, stabilization energies, NICS, ρ_{RCP}, and $\nabla^2\rho_{RCP}$ for all the systems. We calculate the stabilization energies as the difference in energy between the molecules reported and the independent fragments such as

$$C_nH_n \rightarrow {}^{n}/_{2}C_2H_2 \text{ (molecular)}$$

These reported energies are corrected for thermal parameters

$$B_nN_nH_{2n} \rightarrow n\text{BNH}_2 \text{ (molecular)}$$

$$N_nP_nF_{2n} \rightarrow n\text{NPF}_2 \text{ (molecular)}$$

$$(HX)_n \rightarrow n\text{HX (supramolecular)}$$

(zero-point energies and the entropy corrections). Note that the stabilization energies for the weakly interacting systems are corrected for basis set superposition errors using counterpoise corrections.[25]

We first discuss the magnetic criteria for characterizing aromaticity/antiaromaticity in these systems. As evident from Table 1, all the systems except C_4H_4 show aromaticity (negative NICS). Among the six-membered rings, the aromaticity in the systems follows the order $C_6H_6 > N_3P_3F_6 > B_3N_3H_6$, following the order of decreasing covalency in these systems. For C_6H_6, the conjugation is most effective because of pπ−pπ overlap, while it decreases for $N_3P_3F_6$, because of less-effective pπ−dπ overlap. For borazine, however, such orbital overlap is poor, and the stabilization in $B_3N_3H_6$ is primarily due to CT from N to B. In the four-membered systems, aromaticity follows the order $N_2P_2F_4 > B_2N_2H_4 > C_4H_4$ (antiaromatic). The decrease in aromaticity from $N_2P_2F_4$ to $B_2N_2H_4$ arises as a result of stronger covalency in the N−P bond compared to that of the B−N bond.

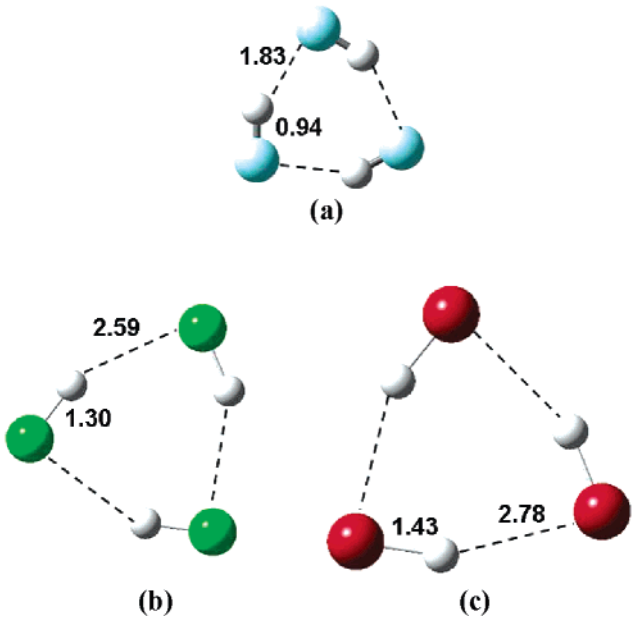


Figure 3. Ground-state optimized geometries of weakly interacting systems (a) $(HF)_3$, (b) $(HCl)_3$, and (c) $(HBr)_3$. Bond lengths in Å are shown for each structure.

In Figure 3, we report the ground-state optimized geometries of the weakly interacting systems: $(HX)_3$, (X = F, Cl, and Br). $(HF)_3$ forms the most compact cyclic structure (as evident from the stabilization energies and BLA), followed by $(HCl)_3$ and $(HBr)_3$. All three of these H-bonded systems have a stable ground-state cyclic geometry, as evident from the absence of any imaginary frequencies in their optimized structures (see Supporting Information). Thus, there is a predominating tendency for these systems to assume a cyclic geometry. This is, in principle, identical to the origin of the high symmetric 0 BLA structures of the conventional aromatic systems such as benzene. The existence of aromaticity is evident from the NICS values (in ppm) of −2.94, −1.98, and −1.89 for $(HF)_3$, $(HCl)_3$, and $(HBr)_3$, respectively. The decreasing aromaticity in the series follows the trend of their decreasing strength of H bonding and the stability of the cyclic H-bonded systems.

The issue of aromaticity in H-bonded systems has been dealt with in the literature in the context of resonance-assisted H bonding for π-conjugated systems such as the enol form of β-diketone.[26] For example, the *cis*-2-enol form of acetylacetone, where the proton is shared equally by the two O atoms, corresponds to the ground-state geometry.[27] The stability of these structures is understood on the basis of the formation of a six-membered ring containing 6π electrons and, thus, aromatic characteristics. However, note that, for our $(HX)_3$ systems, the stability has its origins in the delocalization of the σ electrons. The stabilization energies (after incorporation of zero-point and entropy corrections) associated with such σ aromaticity in $(HF)_3$, $(HCl)_3$, and $(HBr)_3$ are −10.94, −4.23, and −1.67 kcal/mol, respectively.

For a quantitative estimation of the role of H bonding in introducing polarization across the full perimeter of these cyclic systems, we calculate the polarizabilities for these systems as $\alpha_{ring} = \bar{\alpha}_{trimer} - 3\bar{\alpha}_{monomer}$, where we define the

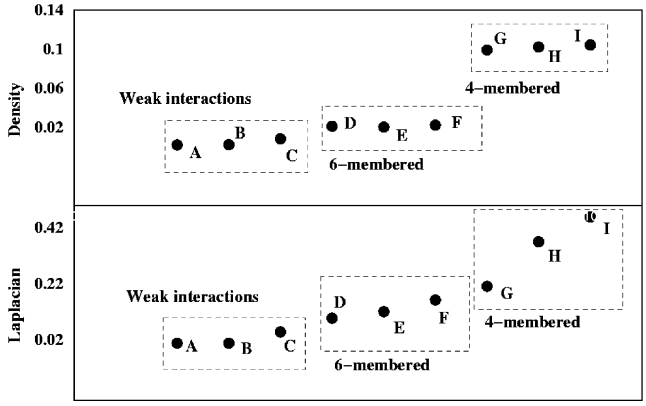


Figure 4. Charge density (upper panel) and laplacian of the charge density (lower panel) at the ring critical points of (A) $(HCl)_3$, (B) $(HBr)_3$, (C) $(HF)_3$, (D) $B_3N_3H_6$, (E) $N_3P_3F_6$, (F) C_6H_6, (G) $B_2N_2H_4$, (H) C_4H_4, and (I) $N_2P_2F_4$.

isotropic average polarizability for the trimer and monomer as

$$\bar{\alpha} = \frac{1}{3}\sum_i (\alpha_{ii}) \qquad (1)$$

where the sums are over the coordinates *x*, *y*, and *z* (*i* = *x*, *y*, and *z*). The calculated polarizabilities are $\bar{\alpha}_{ring}(HF) = +2.24$ au, $\bar{\alpha}_{ring}(HCl) = +7.97$ au, and $\bar{\alpha}_{ring}(HBr) = +12.16$ au. Note that, for all three of the H-bonded complexes, there is a cooperative enhancement of polarizability, suggesting extended delocalization across the ring. Also, the order of increasing polarizability, $\bar{\alpha}_{ring}(HF) < \bar{\alpha}_{ring}(HCl) < \bar{\alpha}_{ring}(HBr)$, follows the decreasing electronegativity in X along group 17 of the periodic table. This leads to a more facile delocalization of σ electrons for the weaker H-bonded systems as compared to the strongest H bonding in HF. However, the aromaticity index (NICS) suggests larger aromaticity for HF and HCl compared to HBr primarily because of a more compact structure (smaller surface area), leading to stronger diamagnetic ring current.

An analysis of the charge density (ρ_{RCP}) and the Laplacian of the charge density ($\nabla^2\rho_{RCP}$) at the ring critical points for these systems reveals clear distinctions between the nature of interactions in the rings (Figure 4). Both ρ_{RCP} and $\nabla^2\rho_{RCP}$ show maximum localizations for the four-membered rings C_4H_4, $B_2N_2H_4$, and $N_2P_2F_4$, followed by the six-membered rings C_6H_6, $B_3N_3H_6$, and $N_3P_3F_6$ (see Table 1 for the values for each system). The H-bonded systems also show a localization of electrons at the ring critical points, suggesting substantial stability in these cyclic systems, supporting results derived from our NICS calculations. Consistent with the maximum stability of the $(HF)_3$ H-bonded system, both ρ_{RCP} and $\nabla^2\rho_{RCP}$ are also highest for it. Thus, both NICS and topological aspects suggest substantial electronic delocalizations across the weakly interacting rings.

IV. $\sigma-\pi$ Separation Analysis

As already discussed, the delocalization of the π electrons over the cyclic architectures differ for the homoatomic and heteroatomic systems. Unlike carbon, nitrogen and phosphorus do not have a straightforward $\sigma-\pi$ separation of their

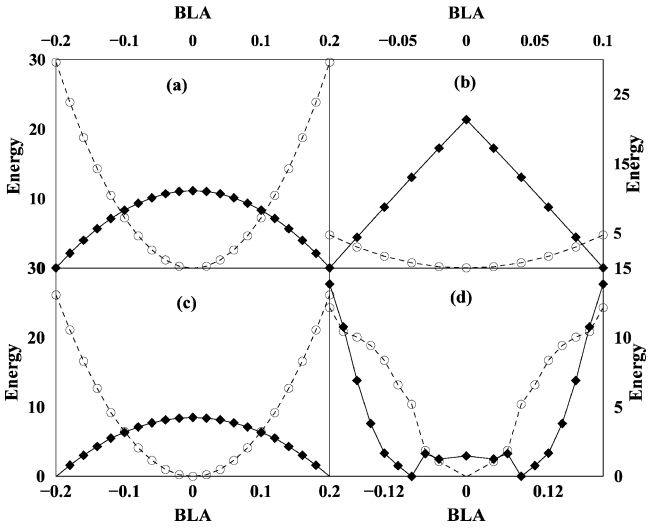


Figure 5. $\sigma-\pi$ separation of energies for (a) C_6H_6, (b) C_4H_4, (c) $B_3N_3H_6$, and (d) $N_3P_3F_6$. Energies are reported in kcal/mol and BLA in Å. Circles and squares correspond to σ and π energies, respectively.

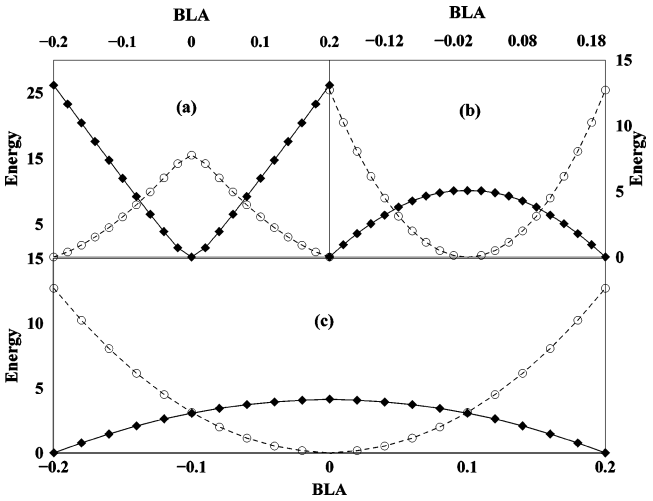


Figure 6. $\sigma-\pi$ separation of energies for (a) $N_2P_2F_4$, (b) $B_2N_2H_4$ (planar), and (c) $B_2N_2H_4$ (puckered). Energies are reported in kcal/mol and BLA in Å. Circles and squares correspond to σ and π energies, respectively.

lower energy levels. Currently, there are efforts to understand aromaticity on the basis of the all-electron models where both σ and π electrons are explicitly taken into account and the overall structure is controlled by the predominance of either of the energy scales.[28] One of the most direct methods for considering the role of σ and π electrons is to separate the total energy of the system into σ and π components. For realizing the σ contribution to the structure, we consider the highest spin (H. S.) state for the systems and freeze all the π electrons in H. S. configuration. Thus, for the six-membered rings such as C_6H_6, $B_3N_3H_6$, and $N_3P_3F_6$, the H. S. state corresponds to $S = 3$, while for the four-membered systems such as C_4H_4, $B_2N_2H_4$, and $N_2P_2F_4$, the H. S. state has a spin of $S = 2$. The π energy for a system is calculated as $E(\pi) = E(\text{G. S.}) - E(\text{H. S.})$, where E(G. S.) corresponds to the energy of the singlet ($S = 0$) state. We have recently benchmarked this method of $\sigma-\pi$ separation for both organic and inorganic molecules.[29]

In Figure 5, we report this $\sigma-\pi$ analysis for C_6H_6, C_4H_4, $B_3N_3H_6$, and $N_3P_3F_6$, as a function of distortion (BLA) in the rings using $\Delta E(\pi) = \Delta E(\text{G. S.}) - \Delta E(\text{H. S.})$, where the energies are scaled so that the most stable structure corresponds to the zero of energy. Note that we define ΔE(H. S.) $= \Delta E(\sigma)$. For benzene (Figure 5a), the symmetric D_{6h} structure (0 BLA) is associated with the stabilization of the σ energy, while the π energy stabilizes the distorted structure. The energy scale for σ equalization overwhelms the π distortion (by 20 kcal/mol), and thus, the symmetric structure for benzene is stabilized. One can clearly observe the role of the σ energies in controlling the structure of benzene. Similar results have also been reported previously.[30] Contrary to the situation for benzene, C_4H_4 (Figure 5b) shows π distortion overwhelming σ equalization. Thus, the distorted D_{2h} structure is stabilized over the undistorted structure. Note that, for these homoatomic systems, we derive results identical to those well-known from π-only electron theories claiming benzene to be aromatic (0 BLA) and C_4H_4 to be JT-distorted antiaromatic.

The heteroatomic B−N and P−N systems also show similar electronic features (Figure 5c and d). $B_3N_3H_6$ is identical to benzene in being σ-equalized and π-distorted. However, compared to benzene, the σ-equalization energy is smaller (by 5 kcal/mol), suggesting $B_3N_3H_6$ to be less aromatic, a result already derived from both NICS and topological analysis. $N_3P_3F_6$, on the contrary, shows double equalization, and both π and σ energies stabilize the 0 BLA structure. While, σ equalization is expected for a cyclic structure, π equalization suggests the predominating p(π)−d(π) delocalizations.

$N_2P_2F_4$ (Figure 6a), on the other hand, is σ-distorted but π-equalized (with π equalization $> \sigma$ distortion by 10 kcal/mol), suggesting strong π delocalization overwhelming minor JT distortion. Thus, $N_2P_2F_4$ may be considered as π-aromatic. In $B_2N_2H_4$, for both the planar (Figure 6b) and the puckered structures (Figure 6c), σ equalization overwhelms π distortion (by 10 kcal/mol). Thus, both the structures correspond to predominantly σ-aromatic 0 BLA geometries.

From the above $\sigma-\pi$ analysis, it is clear that JT distortion in the backbones leads to structures with large BLAs. We have performed an analysis of the fragmentation of the total energy into contributions from the nuclear−nuclear (V_{nn}), electron−nuclear (V_{en}), electron−electron (V_{ee}), and kinetic energy (K.E) components as a function of BLA. The results for each of the systems are shown in Figure 7. For all cases, the electron−nuclear (V_{en}) component favors distortion, while V_{nn}, V_{ee}, and K.E have a preference for the undistorted structure. The V_{nn}, V_{ee}, and K.E components are stabilized in structures with 0 BLA as they are associated with a complete delocalization of electrons across the ring. For large BLAs, electrons are localized in the shorter bonds. Thus, the actual preference for the highly symmetric or distorted structure is governed by the competition between all other components and V_{en}. In C_6H_6, V_{en} is overwhelmed by the other components (Figure 7a), while in the case of C_4H_4, V_{en} is the major component (Figure 7b) and the structure is overall distorted. The preference for the heteroatomic systems such as $B_3N_3H_6$ (Figure 7c), $B_2N_2H_4$ (planar) (Figure 7d),

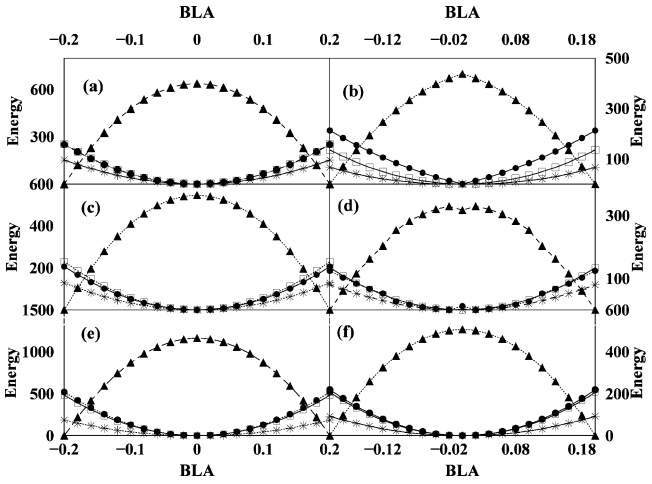


Figure 7. Variation of individual energy components (in kcal/mol) V_{ee} (circles), V_{nn} (squares), V_{ne} (triangles), and V_{KE} (stars) with BLA in Å for (a) C_6H_6, (b) C_4H_4, (c) $B_3N_3H_6$, (d) $B_2N_2H_4$ (planar), (e) $N_3P_3F_6$, and (f) $N_2P_2F_4$.

$N_3P_3F_6$ (Figure 7e), and $N_2P_2F_4$ (Figure 7f) adapting to a highly symmetric structure (0 BLA) is also clearly understood from the fact that, for these systems, V_{en} is only a minor component.

V. Conclusions

We have considered aromaticity and antiaromaticity in various molecules. Organic molecules such as C_6H_6 and C_4H_4 are stabilized through isotropic delocalization of the π electrons over the full perimeter of the rings. The CT and $p(\pi)-d(\pi)$ interactions in $B_3N_3H_6$ and $N_3P_3F_6$, respectively, lead to aromaticity in these systems although the aromatic character is less than that of benzene. Four-membered heteroatomic systems such as $N_2P_2F_4$ and $B_2N_2H_4$ are also aromatic.

Apart from the covalently bonded systems, the weakly interacting H-bonded systems also have aromatic characteristics. In fact, it is the weak aromaticity developed as a result of the nonlocal nature of these interactions that stabilizes such systems. Finally, we propose that aromaticity is a single parameter that includes all specific interactions in the weakly interacting cyclic systems and provides a global tool to understand their structures.

Acknowledgment. S.S.M. thanks CSIR for the research fellowship. S.K.P. acknowledges CSIR and DST, Government of India, for the research grants.

Supporting Information Available: Cartesian coordinates, total energy (in Hartrees), frequency calculations for all the structures, and complete reference 19. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) (a) Minkin, V. I.; Glukhovtsev, M. N.; Simkin, B. Y. *Aromaticity and Antiaromaticity*; Wiley: New York, 1994. (b) Shaik, S.; Shurki, A.; Danovich, D.; Hilberty, P. C. *Chem. Rev.* **2001**, *101*, 1501−1539. (c) Gomes, J. A. N. F.; Mallion, R. B. *Chem. Rev.* **2001**, *101*, 1349−1383.

(2) (a) Mesbah, W.; Prasang, C.; Hofmann, M.; Geiseler, G.; Massa, W.; Berndt, A. *Angew. Chem., Int. Ed.* **2003**, *42*, 1717−1719. (b) Schiemenz, B.; Huttner, G. *Angew. Chem., Int. Ed. Engl.* **1993**, *32*, 297.

(3) (a) Takanashi, K.; Lee, V. Y.; Matsuno, T.; Ichinohe, M.; Sekiguchi, A. *J. Am. Chem. Soc.* **2005**, *127*, 5768−5769. (b) Lee, V. Y.; Matsuno, T.; Ichinohe, M.; Sekiguchi, A. *J. Am. Chem. Soc.* **2004**, *126*, 4758−4759.

(4) (a) Li, X.; Kuznetsov, A.; Zhang, H.-F.; Boldyrev, A. I.; Wang, L. *Science* **2001**, *291*, 859−861. (b) Li, X.; Zhang, H.-F.; Wang, L.-S.; Kuznetsov, A. E.; Cannon, N. A.; Boldyrev, A. I. *Angew. Chem., Int. Ed.* **2001**, *40*, 1867−1870. (c) Kuznetsov, A.; Boldyrev, A. I.; Li, X.; Wang, L.-S. *J. Am. Chem. Soc.* **2001**, *123*, 8825−8831.

(5) (a) Kuznetsov, A. E.; Birch, K.; Boldyrev, A. I.; Li, X.; Zhai, H.; Wang, L. *Science* **2003**, *300*, 622−625. (b) Chen, Z.; Corminboeuf, C.; Heine, T.; Bohmann, J.; Schleyer, P. v. R. *J. Am. Chem. Soc.* **2003**, *125*, 13930−13931.

(6) (a) Wannere, C. S.; Corminboeuf, C.; Wang, Z.-X.; Wodrich, M. D.; King, R. B.; Schleyer, P. v. R. **2005**, *127*, 5701−5705. (b) Santos, J. C.; Tiznado, W.; Contreras, R.; Fuentealba, P. *J. Chem. Phys.* **2004**, *120*, 1670. (c) Santos, J. C.; Andres, J.; Aizman, A.; Fuentealba, P. *J. Chem. Theory Comput.* **2005**, *1*, 83.

(7) (a) Heilbronner, E. *Tetrahedron Lett.* **1964**, *29*, 1923−1928. (b) Ajami, D.; Oeckler, O.; Simon, A.; Herges, R. *Nature* **2003**, *426*, 819−821. (c) Kawase, T.; Oda, M. *Angew. Chem., Int. Ed.* **2004**, *43*, 4396−4398.

(8) (a) Jutzi, P.; Mix, A.; Rummel, B.; Schoeller, W. W.; Neumann, B.; Stammler, H.-G. *Science* **2004**, *305*, 849. (b) Datta, A.; Pati, S. K. *J. Am. Chem. Soc.* **2005**, *127*, 3496−3500.

(9) (a) Rio, D. D.; Galindo, A.; Resa, I.; Carmona, E. *Angew. Chem., Int. Ed.* **2005**, *44*, 1244−1247. (b) Resa, I.; Carmona, E.; Gutierrez-Puebla, E.; Monge, A. *Science*, **2004**, *305*, 1136. (c) Xie, Y.; Schaefer, H. F., III; King, R. B. *J. Am. Chem. Soc.* **2005**, *127*, 2818−2819.

(10) March, J. *Advanced Organic Chemistry: Reactions, Mechanisms and Structure*, 4th edition; John Wiley and Sons: New York, 1992.

(11) (a) Allcock, H. R. *Chem. Rev.* **1972**, *72*, 315. (b) Kiran, B.; Phukan, A. K.; Jemmis, E. D. *Inorg. Chem.* **2001**, *40*, 3615−3618. (c) Boyd, R. J.; Choi, S. C.; Hale, C. C. *Chem. Phys. Lett.* **1984**, *112*, 136. (d) Jemmis, E. D.; Kiran, B. *Inorg. Chem.* **1998**, *37*, 2110−2116. (e) Krishnamurthy, S. S.; Sau, A. C.; Woods, M. *Adv. Inorg. Chem. Radiochem.* **1978**, *21*, 41−112. (f) Soncini, A.; Domene, C.; Engelberts, C. C.; Fowler, P. W.; Rassat, A.; Lenthe, J. H.; Havenith, R. W. A.; Jenneskens, L. W. *Chem. Eur. J.* **2005**, *11*, 1257−1266.

(12) (a) Baceiredo, A.; Bertrand, G.; Majoral, J.-P.; Sicard, G.; Juad, J.; Galy, J. *J. Am. Chem. Soc.* **1984**, *106*, 6088−6089.

(13) (a) Stone, A. J. *The Theory of Intermolecular Forces*; Oxford University Press: New York, 1996. (b) Ratajczak, H.; Orville-Thomas, W. J. *Molecular Interactions*; John Wiley and Sons: New York, 1980.

(14) (a) Steiner, T. *Angew. Chem., Int. Ed. Engl.* **2002**, *41*, 48. (b) Desiraju, G. R.; Steiner, T. *The Weak Hydrogen Bond in Structural Chemistry and Biology*; Oxford University Press: New York, 1999.

(15) (a) Etter, M. C. *Acc. Chem. Res.* **1990**, *23*, 120. (b) Etter, M. C.; MacDonald, J. C.; Bernstein, J. *Acta Crystallogr., Sect. B* **1990**, *46*, 256.

(16) (a) Radhakrishnan, T. P.; Herndon, W. C. *J. Phys. Chem.* **1991**, *95*, 10609. (b) Bernstein, J.; Davis, R. E.; Shimoni, L.; Chang, N.-L. *Angew. Chem., Int. Ed. Engl.* **1995**, *34*, 1555.

(17) Scheiner, S. *Hydrogen Bonding: A Theoretical Perspective*; Oxford University Press: New York, 1997.

(18) (a) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372. (b) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.

(19) *Gaussian 03*, revision B.05; Gaussian, Inc.: Pittsburgh, PA, 2003.

(20) *Conquest*, version 1.7; Cambridge Crystallographic Database, Cambridge Crystallographic Data Centre: Cambridge, U. K., 2004.

(21) Paetzold, P.; Richter, A.; Thijssen, T.; Wurtenberg, S. *Chem. Ber.* **1979**, *112*, 3811.

(22) Paetzold, P.; von Plotho, C.; Schmid, G.; Boese, R.; Schrader, B.; Bougeard, D.; Pfeiffer, U.; Gleiter, R.; Schafer, W. *Chem. Ber.* **1984**, *117*, 1089.

(23) Schleyer, P. v. R.; Maerker, C.; Dransfeld, A.; Jiao, H.; Eikema-Hommas, N. J. R. v. *J. Am. Chem. Soc.* **1996**, *118*, 6317.

(24) (a) Bader, R. F. W. *Atoms in Molecules: A Quantum Theory*; Clarendon Press: Oxford, U. K., 1995. (b) Cole, J. M.; Copley, R. C. B.; McInyre, G. J.; Howard, J. A. K.; Szablewski, M.; Cross, G. H. *Phys. Rev. B* **2002**, *65*, 125107−(1−11). (c) Ranganathan, A.; Kulkarni, G. U. *Proc. Indian Acad. Sci., Chem. Sci.* **2003**, *115*, 637−647. (d) Luana, V.; Pendas, A. M.; Costales, A.; Carriedo, G. A.; G-Alonson, F. J. *J. Phys. Chem. A* **2001**, *105*, 5280−5291.

(25) (a) Hobza, P.; Zahradnik, R. *Chem. Rev.* **1988**, *88*, 871. (b) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553.

(26) (a) Gilli, G.; Bellucci, F.; Ferretti, V.; Bertolasi, V. *J. Am. Chem. Soc.* **1989**, *111*, 1023. (b) Mohajeri, A. *THEOCHEM*, **2004**, *678*, 201. (c) Bertolasi, V.; Gilli, P.; Ferretti, V.; Gilli, G. *J. Am. Chem. Soc.* **1991**, *113*, 4917. (d) Wojtulewski, S.; Grabowski, S. J. *THEOCHEM*, **2003**, *621*, 285.

(27) (a) Dannenberg, J. J.; Rios, R. *J. Phys. Chem.* **1994**, *98*, 6714. (b) Craw, J. S.; Bacskay, G. B. *J. Chem. Soc., Faraday Trans.* **1992**, *88*, 2315. (c) Gilli, P.; Bertolasi, V.; Ferretti, V.; Gilli, G. *J. Am. Chem. Soc.* **1994**, *116*, 909.

(28) (a) Jug, K.; Hilberty, P. C.; Shaik, S. *Chem. Rev.* **2001**, *101*, 1477. (b) Jug, K.; Koster, A. M. *J. Am. Chem. Soc.* **1990**, *112*, 6772.

(29) Datta, A.; Pati, S. K. *J. Chem. Theory Comput.* **2005**, *1*, 824.

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

CT0501598

JCTC Journal of Chemical Theory and Computation

Relativistic DFT Calculation of ^{119}Sn Chemical Shifts and Coupling Constants in Tin Compounds

Alessandro Bagno,*,† Girolamo Casella,‡ and Giacomo Saielli§

Dipartimento di Scienze Chimiche, Università di Padova, via Marzolo, 1-35131 Padova, Italy, Dipartimento di Chimica Inorganica e Analitica "Stanislao Cannizzaro", Università di Palermo, Viale delle Scienze Parco d'Orleans II, 90128 Palermo, Italy, and Istituto CNR per la Tecnologia delle Membrane, Sezione di Padova, via Marzolo, 1-35131 Padova, Italy

Received July 18, 2005

Abstract: The nuclear shielding and spin−spin coupling constants of ^{119}Sn in stannane, tetramethylstannane, methyltin halides $Me_{4-n}SnX_n$ (X = Cl, Br, I; n = 1−3), tin halides, and some stannyl cations have been investigated computationally by DFT methods and Slater all-electron basis sets, including relativistic effects by means of the zeroth order regular approximation (ZORA) method up to spin−orbit coupling. Calculated ^{119}Sn chemical shifts generally correlate well with experimental values, except when several heavy halogen atoms, especially iodine, are bound to tin. In such cases, calculated chemical shifts are almost constant at the scalar (spin-free) ZORA level; only at the spin−orbit level is a good correlation, which holds for all compounds examined, attained. A remarkable "heavy-atom effect", analogous to that observed for analogous alkyl halides, is evident. The chemical shift of the putative stannyl cation (SnH_3^+) has also been examined, and it is concluded that the spectrum of the species obtained in superacids is inconsistent with a simple SnH_3^+ structure; strong coordination to even weak nucleophiles such as FSO_3H leads to a very satisfactory agreement. On the contrary, the calculated ^{119}Sn chemical shift of the trimesitylstannyl cation is in very good agreement with the experimental value. Coupling constants between ^{119}Sn and halogen nuclei are also well-modeled in general (taking into account the large uncertainties in the experimental values); relativistic spin−orbit effects are again quite evident. Couplings to ^{13}C and ^{1}H also fall, on the average, on the same correlation line, but individual values show a significant deviation from the expected unit slope.

Introduction

The chemistry of tin compounds is important in a variety of contexts, spanning basic research and industrial applications.[1−3] Tin exhibits two oxidation states, Sn(II) and Sn(IV), the latter being the more stable. Organometallic derivatives of Sn(IV) are produced in bulk amounts for a large variety of industrial, agricultural, and biological uses.[4,5] Their use in human cancer treatment is also documented.[6,7]

Most of the structural properties of Sn(IV) compounds arise from its ability to expand its coordination number; this is often higher than the expected four, particularly when bound to more electronegative atoms or to weak donor ligands. This ability is responsible for differences between the solution phase and the solid-state structure of the same compound.

Although Mössbauer spectroscopy is a well-established technique for structure investigations of tin compounds in the solid state and frozen solutions, tin NMR is a more

* Corresponding author fax: +39 0498275239; e-mail: alessandro.bagno@unipd.it.
† Università di Padova.
‡ Università di Palermo.
§ Istituto CNR per la Tecnologia delle Membrane.

10.1021/ct050173k CCC: $33.50
Published on Web 11/11/2005

generally applicable tool to probe their structure and reactivity in solution. Natural tin occurs as three magnetically active isotopes: ^{115}Sn (natural abundance = 0.35%), ^{117}Sn (natural abundance = 7.61%), and ^{119}Sn (natural abundance = 8.58%).[1] Owing to their fairly high natural abundance, the ^{117}Sn and ^{119}Sn isotopes are amenable to experimental NMR studies, although ^{119}Sn is generally preferred, owing to its higher magnetogyric ratio. Very few usages of ^{115}Sn in NMR are reported.[8]

^{119}Sn chemical shifts cover a range from ca. +4000 to −2500 ppm, using tetramethyltin ($SnMe_4$) as a reference, and several reviews on tin NMR have been published.[9−11] As is the case for most metal nuclei, there are few general rules to predict the relationship between structure and NMR spectral features. It is now established that several contributions affect ^{119}Sn chemical shifts, such as the nature of the ligands, the coordination number, the interaction with the solvent, the temperature, and the occurrence of self-association processes or inter-/intramolecular interactions. For example, for Sn(IV), an approximate correlation between the ^{119}Sn chemical shift and its coordination environment is observed: increasing the coordination number causes an increased shielding. Therefore, from the value of the chemical shift, it is possible to estimate the coordination number.[9−11] However, a relatively high correlation can only be observed for analogous compounds, for example, for organotin(IV) compounds with carbohydrate derivatives.[12,13]

A survey about coupling constants between tin and several nuclei has been published, and the magnitude and sign of these couplings is often useful in structural investigations.[14] When another nucleus is coupled to tin, normally both ^{119}Sn and ^{117}Sn satellite peaks are observed. $^1J(^{119}\text{Sn},^{13}\text{C})$ and $^2J(^{119}\text{Sn},^1\text{H})$ couplings obtained from the satellite signals are very useful in the structural determination of organotin(IV) compounds, and some empirical equations have been proposed to relate the spin−spin coupling constants to the C−Sn−C angle in dialkyltin(IV) derivatives.[15] Moreover, a Karplus-like dependence of $^3J(^{119}\text{Sn},^2\text{H})$ has been observed.[16]

Other common Sn(IV) compounds, widely used in synthesis, are tin halides. For these, the direct measurement of $^1J(^{119}\text{Sn,X})$ (X = Cl, Br, I) is hampered by the fact that the extremely short T_1 of the quadrupolar halogen nuclei (generally < 1 μs) leads, at most, to a broadening of the tin NMR signal through scalar relaxation of the first kind. Thus, these direct spin−spin coupling constants have been derived by means of relaxation studies,[17−19] although some direct measurements of $J(^{119}\text{Sn},^{35}\text{Cl})$ coupling constants in organotin chlorides have been reported.[20]

The calculation of NMR properties by means of quantum-chemical methods is becoming an increasingly important tool in NMR spectroscopy. There is a substantial and growing data and knowledge base, indicating that, when suitable methods are adopted, all relevant molecular properties (nuclear shielding and spin−spin coupling constants) can be predicted with outstanding accuracy.[21a] Like in the case of other nuclei, the theoretical study of chemical shifts and coupling constants of ^{119}Sn can usefully complement the structural analysis performed by means of experimental NMR data.

One of the issues associated with heavy-atom nuclei is the importance of relativistic effects thereon. Ziegler and co-workers have carried out a number of such pioneering investigations (of ^{183}W, ^{195}Pt, ^{199}Hg, ^{205}Tl, ^{207}Pb, and ^{235}U) and observed large relativistic effects on their NMR properties.[21b,c] For lighter atoms such as Sn, these effects still have to be investigated in detail. For nuclei of similar atomic number such as Ru,[22] Rh,[23] and Xe[24] and even higher such as W,[25] relativistic effects are intrinsically important (in that they affect nuclear shieldings) but often do not substantially affect the quality of calculated chemical shifts because the latter are the difference between the shieldings in two species, so that some contributions are almost constant and partly cancel. Thus, for example, an excellent agreement between experimental ^{99}Ru chemical shifts and nonrelativistic calculated values was obtained, even if that correlation included complexes where fairly heavy atoms (Sn and I) are bonded to Ru.[26] This cancellation of effects between reference and probe molecules may not hold when only one is subject to strong relativistic effects. This happens when one or more third- or fourth-row atoms (typically iodine) are bonded to an observed light nucleus and the bond has a high s character. In this case, spin−orbit (SO) coupling makes a large contribution to the overall shielding and generally causes the observed nucleus to be unusually shielded (see, e.g., ^{13}C in CI_4, $\delta = -290$ ppm). This effect has been related to a Fermi-contact mechanism intimately connected with the magnitude of the relevant coupling constant.[27]

On the other hand, coupling constants involving heavy-atom nuclei have been found to be subject to large scalar relativistic effects even for moderately heavy nuclei such as those dealt with herein, and even more so for heavier nuclei.[21b,c,28]

^{119}Sn NMR offers a unique environment to further test this situation since the reference compound ($SnMe_4$) only has light atoms, whereas there is a substantial data set pertaining to tin halides, comprising species where one or more halogen atoms from Cl to I are present. Some earlier theoretical investigations by Nakatsuji and co-workers considered Me_nSnH_{4-n} and Me_nSnCl_{4-n} at the self-consistent field level of theory,[29] later corrected by inclusion of SO coupling in the Hamiltonian, to account for the unusual shielding of tin when heavy atoms such as iodine are bound to it.[30] The origin of the SO effect was ascribed by the authors to the Fermi contact term. Semiempirical methods, such as a modified version of the AM1 model Hamiltonian,[31] were also developed to study spin−spin coupling constants, including $^{1,2}J$(Sn,Sn).[32] More recently, relativistic effects have been considered for simple systems such as SnH_4 and $SnMe_4$,[33,34] but for larger organotin derivatives, nonrelativistic DFT methods have been used.[35,36] Generally, a good agreement between theory and experiments has been found for chemical shifts.[37] In contrast, for spin−spin couplings, only highly correlated levels of theory, such as the complete active space self-consistent field[38,39] or relativistic four-component methods,[34] have been able to quantitatively recover the measured ^1J(Sn,H) and ^1J(Sn,C) of SnH_4 and $SnMe_4$. Therefore, it seems that the performance of DFT methods with respect to NMR properties of tin is not yet established

with sufficient generality, especially with regard to Sn compounds containing heavy atoms. It is, therefore, of interest to assess whether DFT is a valuable tool in quantitative predictions of ^{119}Sn NMR properties, particularly spin−spin coupling constants involving it.

A further issue where these calculations may prove useful is connected to the isolation and spectroscopic studies of unstable tin species such as stannyl cations SnR_3^+ and anions SnR_3^-, SnH_3^+ and SnH_3^- being the respective parent compounds.

Computational Details

All calculations have been carried out using DFT as implemented in the Amsterdam density functional (ADF) code,[40] in which frozen-core, as well as all-electron, Slater basis sets are available for all atoms of interest. The ADF code also offers the possibility of taking relativistic effects into account, by means of the two-component zeroth-order regular approximation (ZORA) method,[41] which requires specially optimized basis sets. With each method, it is possible to include either only scalar effects (the ZORA equivalents of Darwin and mass-velocity) or spin−orbit coupling as well.

Our previous works concerning Ru,[22] Rh,[23] Xe,[24] and W[25] compounds, spanning a variety of bonding types and electronic structures, showed that the Becke 88 exchange[42] plus the Perdew 86 correlation[43] (BP) functional performs rather well for the calculation of NMR properties, and this was also selected in this work. Moreover, in one case,[24] we also tested other functionals with no significant differences in the results. The all-electron TZ2P basis set (specially optimized for ZORA calculations) was used with all atoms. Relativistic frozen-core potentials (not to be confused with effective core potential basis sets), required to run relativistic calculations, were generated with the Dirac utility.[40] The geometries were optimized at the BP-ZORA/TZ2P level, taking full advantage of symmetry. All optimized geometries are reported as Supporting Information. Sn−H and Sn−C distances in SnH_4 and $SnMe_4$ were calculated to be 1.715 and 2.184 Å, respectively. The corresponding experimental values are 1.701 and 2.144 Å, respectively.[44] Using the larger QZ4P basis set for the optimization slightly improved the agreement for the Sn−C bond length of $SnMe_4$ (calcd 2.177 Å) but did not affect the Sn−H bond length of SnH_4. As far as the tin−halogen bond distances are concerned, we have found an overestimation of similar magnitude, about 0.05 Å, compared to available experimental values.[44f] Some data are reported as Supporting Information.

The ADF *nmr* property module then allows for the calculation of nuclear shieldings by either method.[45] Shieldings were then calculated at the BP-ZORA/TZ2P scalar and spin−orbit levels. These combinations will be denoted as SC and SO, respectively. In the former case, the isotropic shielding constant σ is given by the sum of dia- and paramagnetic contributions ($\sigma = \sigma_d + \sigma_p$), whereas in the second one, the spin−orbit contribution is also added ($\sigma = \sigma_d + \sigma_p + \sigma_{SO}$). Computed chemical shifts are then determined by the difference of the shielding of the experimental standard $SnMe_4$ ($\delta = 0$) from $\delta = \sigma_{ref} - \sigma$.

Spin−spin coupling constants were calculated with the ADF *cpl* module,[46] with the BP functional and the ZORA method as above. In a nonrelativistic framework, Ramsey's theory[47] dissects the contributions to the coupling constant into the Fermi-contact (FC), diamagnetic spin−orbit (DSO), paramagnetic spin−orbit (PSO), and spin-dipole (SD) terms, so that the reduced coupling constant K is given by $K = K^{FC} + K^{DSO} + K^{PSO} + K^{SD}$. Within the ZORA approximation, the same terms can be calculated, although the FC, SD, and PSO terms contain cross terms with the others. Moreover, if a spin−orbit Hamiltonian is used, the individual FC and SD terms must be evaluated in two independent runs; in this work, we only report the total FC + SD term.

Results

^{119}Sn Chemical Shifts in Alkyltin Halides. In organotin(IV) compounds, the solvent exerts a non-negligible influence on the chemical shift because, as mentioned before, it may strongly coordinate to the metal, thereby causing (among other things) a substantial geometry change. Therefore, to make a consistent comparison between experimental and calculated chemical shifts, we used experimental values acquired in noncoordinating solvents. Taking into account solvent effects would require at least the explicit inclusion of a few solvent molecules and long-range electrostatic contributions,[48] or even the consideration of the full dynamics of the solvated system, as recently done by Bühl et al. for some metal complexes.[49] This would render our computational protocol infeasible for the large set of compounds we have investigated. The experimental chemical shifts are reported in Table 1, together with the results of the calculations discussed below.

For tin halides (SnX_4) and methyltin halides ($Me_{4-n}SnX_n$; X = Cl, Br, I; $n = 1-3$), a strong "heavy-atom" effect is clearly evident: on the basis of the higher electronegativity of Br and I compared to that of C, one would have expected more and more deshielding of the Sn nucleus upon increasing the number of halogen atoms in the series $Me_{4-n}SnX_n$. In contrast, the observed trend is just the opposite, with a large upfield shift which increases strongly as methyl groups are replaced by halogens. This effect, fully analogous to that on ^{13}C, was explained by Nakatsuji and co-workers[30] as originating from spin−orbit coupling in the Hamiltonian, a contribution that becomes more important for atoms of high atomic number. However, the authors did not perform a full relativistic calculation but limited their study to the inclusion of spin−orbit coupling within a Hartree−Fock approach. It is, therefore, of interest to extend such a study by means of a more-detailed relativistic calculation and larger basis set. Moreover, an "experimental" value of the shielding constant of the reference $SnMe_4$ has been reported as $\sigma(SnMe_4) = 2180 \pm 200$ ppm.[50] These data were estimated by means of the experimentally determined ^{119}Sn spin-rotation constant combined with the calculated value of the shielding constant of the free tin atom. The latter calculation did not include relativistic effects.[51] In Table 1, we report the results of our calculations at the two relativistic levels. We note that the scalar relativistic shielding of $SnMe_4$ is very close to the "experimental" value, while the result obtained at the spin−

Table 1. Experimental and Calculated ^{119}Sn Chemical Shifts (ppm)

	ZORA scalar				ZORA spin−orbit						
species	σ_p	σ_d	σ	δ_{calcd}	σ_p	σ_d	σ_{so}	σ	δ_{calcd}	δ_{exptl}	ref
$SnMe_4$	−2747	5030	2283	0	−2772	5032	489	2749	0	0	
$Me_3SnSnMe_3$	−2675	5032	2356	−73	−2700	5034	512	2845	−96	−113	52
SnH_4	−2148	5031	2883	−600	−2165	5033	513	3381	−632	−500[a]	53
SnH_3^-	−1707	5032	3325	−1042	−1724	5034	511	3821	−1072		54
SnH_3^+	−3819	5028	1208	1074	−3847	5030	434	1617	1132	−186	55
$SnH_3^+ \cdot FSO_3H$[b]	−3017	5028	2011	272							
$SnH_3^+ \cdot FSO_3H$[c]	−2861	5028	2167	116							
$SnH_3^+ \cdot 2\ FSO_3H$[b]	−2821	5028	2207	76							
$SnH_3^+ \cdot 2FSO_3H$[c]	−2603	5028	2425	−142							
SnH_3F	−2577	5029	2452	−169							
Mes_3Sn^+[d]	−3589	5032	1443	840	−3624	5034	429	1839	910	806	56
Me_3SnCl	−2933	5031	2098	185	−2963	5034	493	2564	185	164	57
Me_2SnCl_2	−2952	5033	2080	203	−2986	5035	512	2561	188	141.2[e]	58
$MeSnCl_3$	−2860	5034	2175	108	−2895	5036	565	2707	43	21	57
$SnCl_4$	−2731	5036	2306	−23	−2766	5038	687	2960	−210	−150	59
Me_3SnBr	−2945	5030	2085	198	−2975	5033	556	2614	136	128	57
Me_2SnBr_2	−3015	5030	2015	268	−3051	5033	696	2678	71	70	57
$MeSnBr_3$	−2981	5031	2050	233	−3022	5033	999	3010	−261	−165	60
$SnBr_4$	−2877	5031	2154	128	−2919	5034	1609	3723	−973	−638	59
Me_3SnI	−3146	5034	2085	198	−2977	5033	665	2721	28	39	57
Me_2SnI_2	−3086	5032	1946	337	−3132	5035	1037	2939	−190	−159	57
$MeSnI_3$	−3146	5034	1887	395	−3209	5036	1787	3614	−865	−700	61
SnI_4	−3109	5035	1927	356	−3174	5037	3079	4942	−2193	−1701	59
SnI_3Cl	−3024	5035	2012	271	−3080	5038	2527	4485	−1736	−1330, −1347	59
$SnCl_3I$	−2831	5036	2205	78	−2872	5038	1311	3477	−728	−543, −557	59

[a] Extrapolated value. [b] F donor. [c] O donor. [d] Mes = 2,4,6-trimethylphenyl. [e] Saturated in CCl_4.

orbit level is about 600 ppm larger, a deviation almost entirely due to the spin−orbit contribution (σ_{SO}) itself. It is worth noting that σ_{SO} is as large as 500 ppm, even for SnH_4.

Concerning the scalar relativistic results, the diamagnetic contribution to the shielding constant (σ_d) is essentially constant (a 6-ppm variation) through the series, whereas the paramagnetic contribution (σ_p), spanning over 1000 ppm, is quite sensitive to the structure. However, these two contributions alone are not capable of reproducing, even qualitatively, the experimental trend. In fact, chemical shifts calculated at the scalar relativistic level are completely uncorrelated with the experimental values (see Figure 1). In contrast, the spin−orbit contribution is strongly dependent on the number and type of halogen atoms bound to tin; σ_{SO} amounts to 500−600 ppm if tin is coordinated to light atoms or chlorine, 600−1000 ppm for bromine, and 1000−3000 ppm for iodine. The chemical shifts calculated at the ZORA spin−orbit level (Figure 1) are in very good agreement with experimental data; therefore, almost all deviations calculated at the nonrelativistic[36] and scalar relativistic levels can be attributed to the missing σ_{SO} term.

It is worthwhile to discuss the difference between the calculated chemical shift at the scalar and spin−orbit levels in more detail. For the series $Me_{4-n}SnX_n$, upon increasing the number of halogen atoms (i.e., n = 1, 2, 3, 4), this difference respectively amounts to 0, 15, 65, and 187 ppm for X = Cl; 62, 197, 494, and 1101 ppm for X = Br; and 170, 527, 1260, and 2549 ppm for X = I. This trend highlights the importance of spin−orbit coupling when heavy atoms are bound to a central light atom, as already

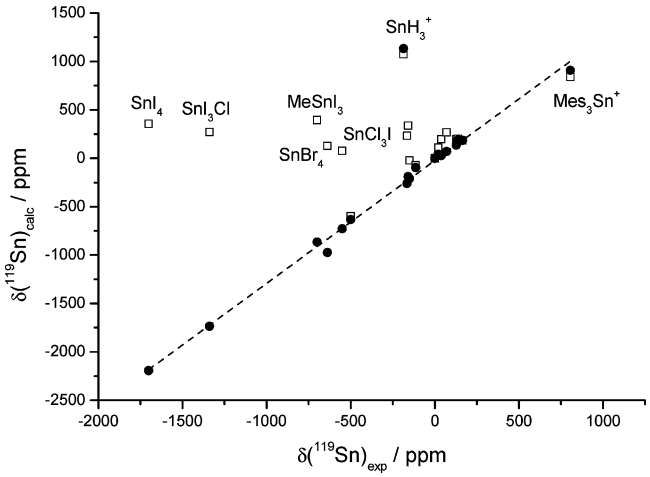


Figure 1. Correlation between calculated and experimental chemical shifts in tin compounds. BP-ZORA scalar (empty squares) and spin−orbit (filled circles), TZ2P basis set. $\delta_{calc} = a + b\delta_{exp}$, $a = -23.0$ ppm, $b = 1.27$; $r = 0.998$. The result for SnH_3^+ is not included in the fit (see text).

mentioned, and the nonadditivity of the effect. We note that even for chlorine such relativistic corrections are not negligible if the number of chlorine atoms is high. A similar, but smaller, effect was found for the ^{13}C nuclei of *o*-bromochlorobenzene.[62]

Finally, we mention that the results of nonrelativistic calculations of ^{119}Sn chemical shifts (see the Supporting Information) are in very good agreement with those obtained at the relativistic ZORA-SC level. As already noted in the Introduction, when no other heavy atoms are directly bound

to the atom of interest, nonrelativistic levels of theory perform rather well for chemical shifts, even for metals. This finding should not obscure the fact that when other heavy atoms are present (as is the case for most species considered herein), relativistic effects must be considered.

^{119}Sn Chemical Shift of Stannyl Cations. NMR has always been concerned with the structure elucidation of unstable species; indeed, the existence of carbocations has been proven by means of this technique. The awareness of this concept has spawned many studies in which the generation of the corresponding electron-deficient species based on Si, Ge, and Sn was attempted. However, despite their formal analogy with carbocations, their existence has been sharply debated, especially in the case of silyl (silicenium) ions R_3Si^+. There is now a general consensus that the stability of silicon, germanium, or tin cations is governed by profoundly different factors than carbocations and that these factors render them extremely electrophilic and incapable of existence as "free" or weakly solvated species in the same sense that is normally attributed to carbocations.[63] Nevertheless, under suitable conditions, silyl cations can be generated.[64]

Quantum chemical calculations, especially of ^{29}Si NMR chemical shifts, have played a major role in establishing these conclusions. The level of accuracy that can currently be attained is such that one can rule out, or raise severe criticism against, structures that do not fit the theoretical expectations and provide indications as to what the actual structures should be. Thus, early experimental ^{29}Si NMR chemical shifts of putative silyl cations (ca. 110 ppm) were deemed too shielded in comparison with the expected values for an isolated silyl cation (ca. 350 ppm). However, it was also shown that coordination with a nucleophile as poor as an argon atom caused substantial shielding from the isolated-ion value, so that care must be taken to compare experimental data, obtained in condensed phases, with appropriate models.[63] As a further example, in our previous work dealing with xenon compounds, we pointed out that the ^{129}Xe spectrum of the species postulated as XeF^+ was inconsistent with that structure and that the bridged $Xe_2F_3^+$ cation would reconcile theoretical and experimental results.[24]

It is then interesting to apply these notions to the case in point. An early attempt at generating a stannyl cation (SnH_3^+) in HSO_3F led to a ^{119}Sn chemical shift of $\delta = -186$ ppm.[55] More recently, Lambert and co-workers reported on the generation of sterically hindered stannyl cations, and ^{119}Sn chemical shifts ranging between 300 and 800 ppm were, thus, observed; in particular, the tris(2,4,6-trimethylphenyl)stannyl cation (Mes_3Sn^+) had $\delta = +806$ ppm.[56] More recently, Lambert was able to obtain the X-ray structure and NMR spectrum of the tris(2,4,6-tri-isopropylphenyl)stannyl cation (Tip_3Sn^+) and provided a computed estimate of its ^{119}Sn chemical shift of +763 ppm, to be compared with the experimental value of +716 ppm.[65]

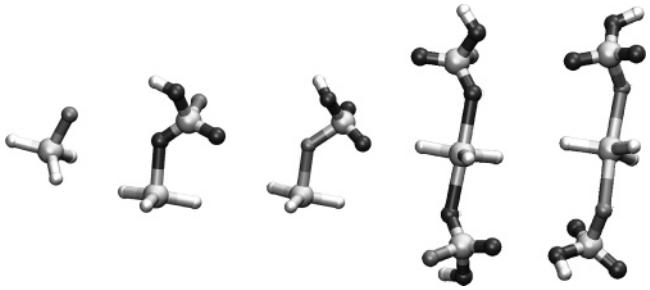


Figure 2. Optimized structures (ZORA scalar/TZ2P) of SnH_3^+·X systems with X = HSO_3F, $2HSO_3F$, and F^-. From left to right: SnH_3F, SnH_3^+·OSO(OH)F, SnH_3^+·FSO_2(OH), SnH_3^+·2OSO(OH)F, and SnH_3^+·$2FSO_2$(OH).

The remarkable agreement provides strong support for the concept that free stannylium ions can be generated in the solid state and in solution. On the other hand, by the same token, it is also evident that many experimental attempts have failed to provide such species. The prototypical example is given by the parent stannylium ion SnH_3^+, for which the experimental[55] chemical shift is some 1300 ppm more shielded than the calculated value of ca. +1100 ppm (Table 1). Therefore, we have strived to provide computed estimates of the chemical shift of relevant stannyl cations in a consistent way and to understand the large variation in experimentally measured values.

The large disagreement indicates that the gas-phase structure of SnH_3^+ is not representative of the actual geometry. Therefore, we have optimized other structures that might have formed in the reaction medium, namely, SnH_3^+·HSO_3F and SnH_3^+·$2HSO_3F$, having oxygen or fluorine as donors, and SnH_3F, as in Figure 2. In Table 2, we report the relevant geometrical parameters.

The calculated ZORA scalar/TZ2P tin chemical shift is strongly influenced by coordination with other species, in full analogy with the behavior of silyl cations. Thus, even a nucleophile as weak as FSO_3H causes a major shielding of the tin nucleus: we obtain $\delta = +116$ ppm and $\delta = -142$ ppm for SnH_3^+·HSO_3F and SnH_3^+·$2HSO_3F$, respectively, with oxygen as the donor, and $\delta = +272$ ppm and $\delta = +76$ ppm for SnH_3^+·HSO_3F and SnH_3^+·$2HSO_3F$, respectively, with fluorine as the donor. Finally, the calculated shift of SnH_3F ($\delta = -169$ ppm), where any cationic character is lost, is in very good agreement with the experimental value of −186 ppm. This could be fortuitous since no evidence of such a compound was reported;[55] nevertheless, it is undeni-

Table 2. Some Geometrical Parameters of Species Related to SnH_3^+

	O donor			F donor		
	r(Sn−O)/Å	*r*(Sn−H)/Å	α/deg[a]	*r*(Sn−F)/Å	*r*(Sn−H)/Å	α/deg[a]
SnH_3^+		1.701	180		1.701	180
SnH_3^+·HSO_3F	2.247	1.699	167	2.240	1.697	162
SnH_3^+·$2HSO_3F$	2.400	1.695	180	2.417	1.696	180
SnH_3F				1.956	1.714	133

[a] Dihedral angle H−Sn−H−H, defining the out-of-plane bending of hydrogen atoms.

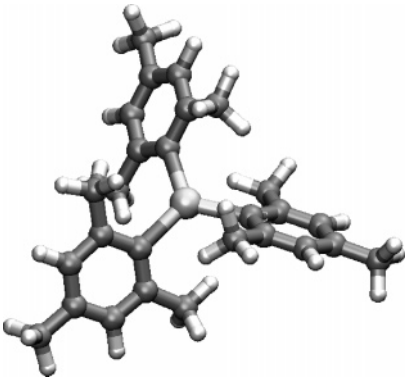


Figure 3. Optimized structure (ZORA scalar/TZ2P) of the tris-(2,4,6-trimethylphenyl)stannylium ion (Mes_3Sn^+).

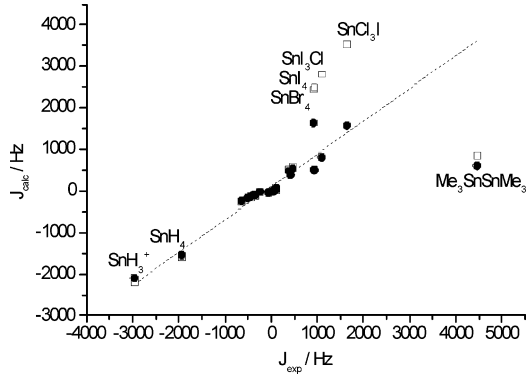


Figure 4. Correlation between calculated and experimental spin−spin coupling constants in tin compounds (ZORA scalar and SO/TZ2P). Scalar relativistic (empty squares) and spin−orbit relativistic (filled circles). $J_{calc} = a + bJ_{exp}$, $a = 101$ Hz, $b = 0.7898$; $r = 0.957$. The result for $Me_3SnSnMe_3$ is not included in the fit (see text).

able that the experimental chemical shift of SnH_3^+ actually pertains to a strongly solvated species; indeed, many compounds where tin is bonded to electron-withdrawing groups resonate in the same range.[56] The chemical shift of $SnH_3^+ \cdot 2OSO(OH)F$ (Figure 2; $\delta = -142$ ppm) is indeed in good agreement with the observed value of −186 ppm. An analogous conclusion was reached by Cremer et al., who computationally investigated SnH_3^+ complexed with one and two water molecules.[66]

On the other hand, the recently reported values[56] of $\delta = +700-800$ ppm for Tip_3Sn^+ and Mes_3Sn^+ agree with the calculated ones and are remarkably close to the value for the naked SnH_3^+ ion. To further probe the scope of ^{119}Sn NMR calculations in a consistent way, we have calculated the ^{119}Sn chemical shift of the trimesitylstannylium ion at the ZORA scalar and SO/TZ2P levels adopted herein. The optimized structure is shown in Figure 3 and features an almost planar coordination geometry of tin (Sn−C−C−C dihedral angle of only 5°).

The ortho methyl groups of the three mesityl substituents, located above and below the coordination plane, prevent tin from interacting with the solvent, in contrast to the case of SnH_3^+. The calculated chemical shift of +840 ppm (Table 1) is, again, in very good agreement with the experimental value (+806 ppm) and is consistent with the tin atom being hardly coordinated to the solvent (benzene).

Spin−Spin Coupling Constants. Before discussing our results in detail, it is of interest to test the performance of the ZORA method in calculating the relativistic contribution to spin−spin coupling constants involving tin. A comparison can be made for the $^1J(^{119}Sn,^1H)$ of SnH_4, for which a four-component random phase approximation approach gave a relativistic effect of about −700 Hz.[34] This method, however, does not properly treat electron correlation: a crude estimate made by the authors to include correlation effects reduced the relativistic contribution to about −550 Hz. The final result was still overestimated (in magnitude) by more than 100 Hz compared to the experimental value. The nonrelativistic value of $^1J(^{119}Sn,^1H)$ that we have obtained at the BP/TZ2P level (−1283 Hz) compared with our ZORA-SC (−1600 Hz) and ZORA-SO (−1550 Hz) results (Table 3) reveals a relativistic contribution of about −300 Hz, in fair agreement with the above proposal, considering the numerous approximations involved.

One-bond tin−halogen coupling constants have been calculated for methyltin and tin halides. These coupling constants cannot be determined by recourse to splittings in the spectra, because the large nuclear quadrupole moments of Cl, Br, and I isotopes cause such signals to have exceedingly short relaxation times and correspondingly broad lines which are normally undetectable; as a consequence, they have been determined indirectly through their effect on the relaxation time of ^{119}Sn (scalar relaxation of the first kind).[17−19] However, this procedure requires some assumption of the rotational correlation time, so there is some uncertainty associated with the experimental values. The coupling constants thus determined are reported in Table 3, together with the results of our calculations.

Spin−spin coupling between tin and chlorine has a non-negligible spin−orbit contribution which increases as the number of iodine atoms bonded to tin increases. In fact, the spin−orbit result improves the calculated $^1J(^{119}Sn,^{35}Cl)$, compared to the scalar calculation, by 6% in $SnCl_4$, 9% in $SnCl_3I$, and 15% in SnI_3Cl. When we consider the couplings with the heavier atoms bromine and iodine, spin−orbit effects become essential in order to obtain a reasonable correlation, as we can see in Figure 4: $^1J(^{119}Sn,^{127}I)$ couplings in SnI_4, SnI_3Cl, and $SnCl_3I$ are all overestimated by some 2000 Hz (a factor of 2−4) compared to the spin−orbit calculation, the latter ones being in much better agreement with the experimentally estimated values (errors being 4−50%). The effect of the number of iodine atoms is, again, not additive. The calculated $^1J(^{119}Sn,^{81}Br)$ in $SnBr_4$ is overestimated, and in this case, the agreement with the experimentally derived data is not quantitative. We note, however, that an unusual dynamical behavior of $SnBr_4$ was reported,[18] which might have affected the accuracy of the estimated coupling constant.

In all cases investigated here, the DSO contribution is negligible (10^{-1} Hz) and not reported in Table 3. In contrast, the PSO contribution is generally important, and it is strongly affected by the inclusion of spin−orbit coupling in the Hamiltonian.

For methyltin halides, we have also calculated the coupling constants with ^{13}C and ^{1}H. The results at the scalar and spin−orbit levels are listed in Table 3. These values are smaller

Table 3. Calculated and Experimental Coupling Constants Involving ^{119}Sn (Hz) in Tin Compounds[a]

		ZORA scalar			ZORA SO				
species	X[b]	PSO	FC + SD	*J* (calcd)	PSO	FC + SD + Cross	*J* (calcd)	*J* (exptl)	ref
$SnMe_4$	C	13.15	−128.97	−115.93	13.70	−118.42	−104.83	−340	67
	H	1.73	6.59	8.45	1.77	4.84	6.74	53.9	68
$Me_3SnSnMe_3$	Sn	−30.53	868.70	838.30	−174.02	775.54	601.66	4460	52
^1J	C	12.65	−43.64	−31.16	12.98	−32.74	−19.92	−240	52
^2J	H	1.91	1.11	2.97	1.94	−0.82	1.07	49	52
^2J	C	0.37	−43.43	−42.99	−0.32	−41.20	−41.46	−56	52
^3J	H	−0.07	−16.04	−15.90	−0.14	−15.70	−15.63	−17.3	52
SnH_4	H	4.74	−1604.34	−1599.63	4.68	−1554.11	−1549.47	(−)1930[c]	69
SnH_3^+	H	19.09	−2207.85	−2188.67	19.12	−2112.89	−2093.68	(−)2960[d]	55
SnH_3^-	H	−1.50	28.18	26.63	−2.90	84.14	81.19	109.4[e]	54
Me_3SnCl	Cl	43.51	247.05	290.53	39.60	237.07	276.64	220[f]	20
	C	18.69	−142.58	−124.02	19.38	−132.10	−112.85	−379.7	70
	H	2.28	3.92	6.29	2.35	2.11	4.54	58.2	71
Me_2SnCl_2	Cl	71.77	309.44	381.18	64.64	299.67	364.28	220[f]	20
	C	21.41	−194.14	−172.88	22.49	−184.73	−162.39	−468.4, −566	65,70
	H	2.29	4.62	6.94	2.40	2.86	5.28	68.2, 68.9	68,72
$MeSnCl_3$	Cl	85.58	393.57	479.10	72.17	384.57	456.70		
	C	20.49	−330.70	−310.38	22.56	−322.90	−300.50		
	H	2.02	15.28	17.26	2.24	13.66	15.86	96.9	72
$SnCl_4$	Cl	83.96	484.37	568.29	59.24	476.32	535.50	470	17
Me_3SnBr	Br	219.81	1036.12	1255.87	94.28	984.96	1079.18		
	C	18.77	−131.81	−113.20	19.09	−121.81	−102.88	−368.9, −380	67,70
	H	2.38	3.27	5.65	2.42	1.45	3.88	57.8	71
Me_2SnBr_2	Br	380.86	1285.94	1666.69	171.95	1231.59	1403.43		
	C	21.63	−166.59	−145.16	21.40	−156.55	−135.35	−442.7	70
	H	2.50	2.23	4.56	2.61	0.29	2.65	66.7	70
$MeSnBr_3$	Br	460.86	1624.97	2085.68	64.62	1577.79	1642.26		
	C	21.15	−269.16	−248.27	19.86	−255.96	−236.35	−640	67
	H	2.31	9.23	11.22	2.33	6.97	8.98		
$SnBr_4$	Br	454.22	1988.92	2442.95	−334.65	1968.18	1633.34	920	18
Me_3SnI	I	217.35	1263.36	1480.67	−131.95	1225.32	1093.32		
	C	18.57	−117.08	−98.68	18.20	−106.43	−88.40		
	H	2.43	2.51	4.88	2.44	0.78	3.17		
Me_2SnI_2	I	390.53	1510.12	1900.57	−210.70	1473.05	1262.27		
	C	21.71	−125.33	−103.85	19.17	−108.35	−89.42		
	H	2.67	−0.32	2.10	2.57	−2.48	−0.15		
$MeSnI_3$	I	481.94	1779.49	2261.31	−648.94	1777.84	1128.78		
	C	22.33	−172.81	−150.76	16.08	−132.27	−116.49		
	H	2.59	0.71	2.84	2.27	−3.29	−1.47		
SnI_4	I	488.65	2009.16	2497.64	−1597.98	2103.98	505.84	940	17
$SnCl_3I$	Cl	84.77	445.64	530.35	42.47	439.83	482.23	378[g]	19
	I	483.80	3052.88	3536.61	−1606.86	3179.70	1572.76	1638	19
SnI_3Cl	Cl	86.08	369.64	455.62	22.96	363.27	386.14	421	19
	I	487.95	2321.11	2808.93	−1628.81	2434.83	805.89	1097	19

[a] BP-ZORA scalar or SO/TZ2P. DSO terms are always negligible and are not reported. [b] X = ^{1}H, ^{13}C, ^{35}Cl, ^{81}Br, ^{119}Sn, and ^{127}I. Coupling constants with ^{1}H in methyl groups are averaged assuming fast rotation. [c] Signs in parentheses have been inferred by comparison with similar molecules. [d] −78 °C. The sign has been assumed equal to the calculated one. [e] −78 °C. [f] Approximate value for triaryltin chlorides.[20] [g] Estimated using T_2 (^{35}Cl) for $SnCl_4$.

than with halogens and, therefore, occupy a small range in the plot of Figure 4. On the whole, they fall into the same correlation line of the other compounds. However, it is of interest to focus on this small region and discuss the behavior of this type of coupling because their magnitude is often related to the coordination pattern of tin; this is presented in Figure 5.

Unexpectedly, even though calculated carbon and proton coupling constants are well correlated with the experimental values, the slope of the linear fit (0.3) is far from unity. Since these couplings are known to be very sensitive to the geometry of coordination around tin (and are commonly employed precisely for this purpose), to check for such an effect for the smaller systems ($SnMe_4$ and SnH_4), we have also repeated the calculation using the larger QZ4P basis set both for geometry optimization and for the calculation of the property. The results, however, were not significantly affected: calculated $^1J(^{119}Sn,^{13}C)$ and $^2J(^{119}Sn,^1H)$ in $SnMe_4$, at the higher level of theory, were −101.6 and +9.4 Hz, respectively, that is, rather similar to the results obtained with

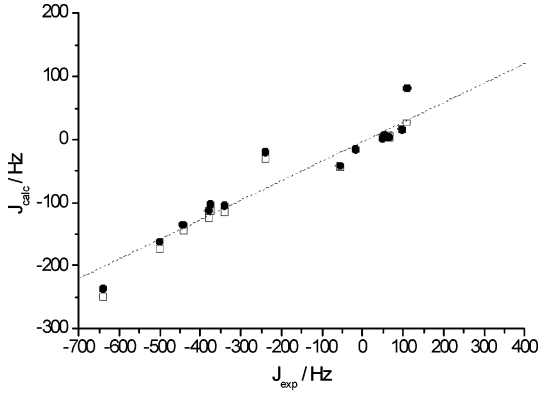


Figure 5. Correlation between calculated and experimental spin−spin coupling constants in tin compounds; expanded view on $J(^{119}Sn,^{1}H)$ and $J(^{119}Sn,^{13}C)$ coupling constants from the data of Figure 4. The fit line (dashed) $J_{calc} = a + bJ_{exp}$ has $a = -3.3$ Hz and $b = 0.309$; $r = 0.957$. Scalar relativistic (empty squares) and spin−orbit relativistic (filled circles).

the TZ2P basis set. On the other hand, calculation (TZ2P basis set) of the coupling constants using a $SnMe_4$ geometry with Sn−C bond lengths constrained to the experimental value yielded some improvement: −146.0 and +9.9 Hz for $^{1}J(^{119}Sn,^{13}C)$ and $^{2}J(^{119}Sn,^{1}H)$, respectively. A further test concerned Me_3SnBr, setting the Sn−Br distance to the experimental value[44f] (see the Supporting Information) while the rest of the molecule was kept at the previously optimized BP/TZ2P geometry (in this case, the calculated Sn−C distance of 2.1713 Å was the same as the experimental value of 2.17 ± 5 Å). Again, some improvement was obtained for $^{1}J(^{119}Sn,^{13}C)$ and $^{2}J(^{119}Sn,^{1}H)$, being −120 and +25 Hz, respectively. However, such values remain quite far from the experimental results.

As a final test, we considered the performance of several GGA functionals. The results are fully reported as Supporting Information (Tables S4 and S5); herein, we will only report on the main conclusions. As we noted in our previous work on xenon compounds,[24] the performance of the various functionals is very similar: for example, $^{1}J(^{119}Sn,^{13}C)$ in $SnMe_4$ (ZORA scalar) ranges from a minimum of −107 Hz with the OPBE functional[73a,73b] to a maximum of −132 Hz with the BLYP functional[73c,73d] against an experimental value of −340 Hz. Therefore, even if BLYP appears to be somewhat superior, it underestimates the experimental result by more than 200 Hz. A slightly better result (−164 Hz at the ZORA scalar level), but still way off the experimental data, is obtained by using the BLYP functional in the calculation of the coupling constant together with the experimental geometry of $SnMe_4$, as discussed above. The same considerations apply to the $^{2}J(^{119}Sn,^{1}H)$ value in $SnMe_4$: the "best" calculated value (BLYP/experimental geometry; about +10 Hz) is less than 20% of the experimental coupling constant. It is presently unclear why the performance is worse than for other similar nuclei; however, such poor performance does not seem to be related to issues such as the choice of functional and basis set, or with geometry effects. We can only note that (a) other groups have reported similar inaccuracies with DFT methods[33,39] and (b), more importantly, there are few if any other examples where minute variations in coupling constants, arising from small structural changes, were investigated.

The only Sn−Sn coupling investigated herein pertains to hexamethylditin, $Me_3SnSnMe_3$. Whereas its calculated ^{119}Sn chemical shift is quite in line with the general level of accuracy attained, some of its coupling constants [most notably $^{1}J(^{119}Sn,^{119}Sn)$ but also $^{1}J(^{119}Sn,^{13}C)$] lie badly off the correlation line. In a pioneering study, experimental values were arrived at indirectly, through a detailed analysis of the ^{1}H and INDOR ^{119}Sn spectrum.[52] Subsequent investigations confirmed the previous data and, at the same time, pointed out the extremely sensitive dependence of such couplings to even minute structural changes.[74] It then appears that ditin species still present a major challenge, in that subtle conformational, steric, and (possibly) solvent effects have to be considered.

We finally comment on the $^{1}J(^{119}Sn,^{1}H)$ values of SnH_3^+ and SnH_3^-. The former calculated value is some 30% off the experimental one, that is, with an error comparable to that of other compounds. Recalling the concerns expressed above on the nature of this species, this fair agreement is probably accidental, and we did not proceed with further evaluations. The value for SnH_3^- is also in rather good agreement with the calculated value. Since, however, no ^{119}Sn data were reported, it is difficult to judge whether the experimental conditions (deprotonation of SnH_4 with sodium in liquid ammonia) really led to SnH_3Na as claimed, although the high polarity of liquid NH_3 may indeed lead to an essentially "free" anion.[54]

Conclusions

The calculation of ^{119}Sn chemical shifts and couplings by means of the ZORA relativistic method yields reliable results that may substantially aid in the structural elucidation of tin compounds. The wide array of species that can be studied includes some where heavy atoms such as iodine are bonded to tin; in such cases, we have shown relativistic spin−orbit corrections to be essential in order to provide a meaningful modeling. We have also shown how such calculations can identify incorrect assignments, like in the case of SnH_3^+. The efficiency of the ADF code in handling these calculations should open the way to their widespread application in a broad range of structural and spectroscopic issues. However, when small variations in coupling constants are sought, like in the case of $J(^{119}Sn,^{1}H)$ and $J(^{119}Sn,^{13}C)$ in alkylstannanes, the performance is poorer despite an ample exploration of possible causes. Hence, there are still important issues to be addressed before such calculations enter into widespread usage.

Supporting Information Available: Cartesian coordinates of all structures optimized, experimental and calculated tin−halogen bond distances of methyltin halides and tin halides in the gas phase, nonrelativistic chemical shifts, and performance tests of various GGA functionals (14 pages). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Davies, A. G. *Organotin Chemistry*; VCH: Weinheim, Germany, 1997.

(2) Blunden, S. J.; Cusack, P. A.; Hill, R. *The Industrial Uses of Tin Chemicals*; The Royal Society of Chemistry: London, 1985.

(3) *Chemistry of Tin*, 2nd ed.; Smith, P. J., Ed.; Blackie Academic & Professional: London, 1998.

(4) Evans, C. J. In *Chemistry of Tin*, 2nd ed.; Smith, P. J., Ed.; Blackie Academic & Professional: London, 1998; Chapter 12, pp 442−479.

(5) Arakawa, Y. In *Chemistry of Tin*, 2nd ed.; Smith, P. J., Ed.; Blackie Academic & Professional: London, 1998; Chapter 10, pp 388−428.

(6) Saxena, A. K.; Huber, F. *Coord. Chem. Rev.* **1989**, *95*, 109−123.

(7) Gielen, M. *Appl. Organomet. Chem.* **2002**, *16*, 481−494.

(8) Meurice, J. C.; Vallier, M.; Ratier, M.; Duboudin, J. G.; Petraud, M. *J. Chem. Soc., Perkin Trans. 2* **1996**, 1311−1313.

(9) Smith, P. J.; Tupčiauskas, A. P. *Annu. Rep. NMR Spectrosc.* **1978**, *8*, 291−370.

(10) Wrackmeyer, B. *Annu. Rep. NMR Spectrosc.* **1985**, *16*, 73−186.

(11) Wrackmeyer, B. *Annu. Rep. NMR Spectrosc.* **1999**, *38*, 203−264.

(12) Grindley, T. B. *Adv. Carbohydr. Chem. Biochem.* **1998**, *53*, 17−142.

(13) Pellerito, L.; Nagy, L. *Coord. Chem. Rev.* **2002**, *224*, 111−150 and references therein.

(14) Wrackmeyer, B. In *Advanced Applications of NMR to Organometallic Chemistry*; Gielen, M., Willem, R., Wrackmeyer, B., Eds.; Wiley: Chichester, U. K., 1996; Chapter 4, pp 87−122.

(15) (a) Lockhart, T. P.; Manders, W. F. *Inorg. Chem.* **1986**, *25*, 892−895. (b) Lockhart, T. P.; Manders, W. F. *J. Am. Chem. Soc.* **1987**, *109*, 7015−7020. (c) Holeček, J.; Lyčka, A. *Inorg. Chim. Acta* **1986**, *118*, L15−L16. (d) Holeček, J.; Nadvornik, M.; Handlir, K.; Lyčka, A. *J. Organomet. Chem.* **1986**, *315*, 299−308.

(16) (a) Quintard, J. P.; Degueil-Castaing, M.; Dumartin, G.; Barbe, B.; Petraud, M. *J. Organomet. Chem.* **1982**, *234*, 27−40. (b) Quintard, J. P.; Degueil-Castaing, M.; Barbe, B.; Petraud, M. *J. Organomet. Chem.* **1982**, *234*, 41−61.

(17) Sharp, R. R. *J. Chem. Phys.* **1972**, *57*, 5321−5330.

(18) Sharp, R. R. *J. Chem. Phys.* **1974**, *60*, 1149−1157.

(19) Sharp, R. R.; Tolan, J. W. *J. Chem. Phys.* **1976**, *65*, 522−530.

(20) Apperley, D. C.; Haiping, B.; Harris, R. K. *Mol. Phys.* **1989**, *68*, 1277−1286.

(21) (a) *Calculation of NMR and EPR Parameters*; Kaupp, M., Bühl, M., Malkin, V. G., Eds.; Wiley-VCH: Weinheim, Germany, 2004. (b) Autschbach, J. In *Calculation of NMR and EPR Parameters*; Kaupp, M., Bühl, M., Malkin, V. G., Eds.; Wiley-VCH: Weinheim, Germany, 2004; Chapter 14, pp 227−247. (c) Autschbach, J.; Ziegler, T. In *Calculation of NMR and EPR Parameters;* Kaupp, M., Bühl, M., Malkin, V. G., Eds.; Wiley-VCH: Weinheim, Germany, 2004; Chapter 15, pp 249−264.

(22) Bagno, A.; Bonchio, M. *Magn. Reson. Chem.* **2004**, *42*, S79−S87.

(23) Orian, L.; Bisello, A.; Santi, S.; Ceccon, A.; Saielli, G. *Chem.—Eur. J.* **2004**, *10*, 4029−4040.

(24) Bagno, A.; Saielli, G. *Chem.—Eur. J.* **2003**, *9*, 1486−1495.

(25) (a) Bagno, A.; Bonchio, M.; Sartorel, A.; Scorrano, G. *ChemPhysChem* **2003**, *4*, 517−519. (b) Bagno, A.; Bonchio, M. *Angew. Chem., Int. Ed.* **2005**, *44*, 2023−2025.

(26) Bühl, M.; Gaemers, S.; Elsevier, C. J. *Chem. Eur. J.* **2000**, *6*, 3272−3280.

(27) (a) Kaupp, M.; Malkina, O. L.; Malkin, V. G.; Pyykkö, P. *Chem. Eur. J.* **1998**, *4*, 118−126. (b) Kaupp, M.; Aubauer, C.; Engelhardt, G.; Klapötke, T. M.; Malkina, O. L. *J. Chem. Phys.* **1999**, *110*, 3897−3902.

(28) Bryce, D. L.; Wasylishen, R. E.; Autschbach, J.; Ziegler, T. *J. Am. Chem. Soc.* **2002**, *124*, 4894−4900.

(29) Nakatsuji, H.; Inoue, T.; Nakao, T. *J. Phys. Chem.* **1992**, *96*, 7953−7958.

(30) Kaneko, H.; Hada, M.; Nakajima, T.; Nakatsuji, H. *Chem. Phys. Lett.* **1996**, *261*, 1−6.

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

(32) (a) Aucar, G. A.; Botek, E.; Gómez, S.; Sproviero, E.; Contreras, R. H. *J. Organomet. Chem.* **1996**, *524*, 1−7. (b) González, J. A.; Aucar, G. A.; Ruiz de Azúa, M. C.; Contreras, R. H. *Int. J. Quantum Chem.* **1997**, *61*, 823−833.

(33) Khandogin, J.; Ziegler, T. *J. Phys. Chem. A* **2000**, *104*, 113−120.

(34) Enevoldsen, T.; Visscher, L.; Saue, T.; Jensen, H. J. A.; Oddershede, J. *J. Chem. Phys.* **2000**, *112*, 3493−3498.

(35) Avalle, P.; Harris, R. K.; Karadakov, P. B.; Wilson, P. J. *Phys. Chem. Chem. Phys.* **2002**, *4*, 5925−5932.

(36) Vivas-Reyes, R.; De Proft, F.; Biesemans, M.; Willem, R.; Geerlings, P. *J. Phys. Chem. A* **2002**, *106*, 2753−2759.

(37) de Dios, A. C. *Magn. Reson. Chem.* **1996**, *34*, 773−776.

(38) Kirpekar, S.; Jensen, H. J. A.; Oddershede, J. *Chem. Phys.* **1994**, *188*, 171−181.

(39) Malkina, O. L.; Salahub, D. R.; Malkin, V. G. *J. Chem. Phys.* **1996**, *105*, 8793−8800.

(40) te Velde, G.; Bickelhaupt, F. M.; Baerends, E. J.; Fonseca Guerra, C.; van Gisbergen, S. J. A.; Snijders, J. G.; Ziegler, T. *J. Comput. Chem.* **2001**, *22*, 931−967. See also: http://www.scm.com.

(41) Jensen, F. *Introduction to Computational Chemistry*; Wiley: Chichester, U. K., 1999; p 204.

(42) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098−3100.

(43) Perdew, J. P. *Phys. Rev. B* **1986**, *33*, 8822−8824.

(44) (a) Vilkov, L. V.; Mastryukov, V. S.; Sadova, N. I. *Determination of the Geometrical Structure of Free Molecules*; Mir Publisher: Moscow, 1983. (b) Wilkinson, G. R.; Wilson, M. K. *J. Chem. Phys.* **1956**, *25*, 784−784. (c) Beagley, B.; McAloon, K.; Freeman, J. M. *Acta Crystallogr. B* **1974**, *30*, 444−449. (d) Clark, H. C.; Furnival, S. G.;

Kwon, J. T. *Can. J. Chem.* **1963**, *41*, 2889−2897. (e) Fujii, H.; Kimura, M. *Bull. Chem. Soc. Jpn.* **1970**, *44*, 2643−2647. (f) Zubieta, J. A.; Zuckerman, J. J. *Prog. Inorg. Chem.* **1978**, *24*, 251−475.

(45) (a) Schreckenbach, G.; Ziegler, T. *J. Phys. Chem.* **1995**, *99*, 606−611. (b) Schreckenbach, G.; Ziegler, T. *Int. J. Quantum Chem.* **1997**, *61*, 899−918. (c) Wolff, S. K.; Ziegler, T. *J. Chem. Phys.* **1998**, *109*, 895−905. (d) Wolff S. K.; Ziegler, T.; van Lenthe, E.; Baerends, E. J. *J. Chem. Phys.* **1999**, *110*, 7689−7698.

(46) (a) Autschbach, J.; Ziegler, T. *J. Chem. Phys.* **2000**, *113*, 936−947. (b) Autschbach, J.; Ziegler, T. *J. Chem. Phys.* **2000**, *113*, 9410−9418.

(47) Ramsey, N. F. *Phys. Rev.* **1953**, *91*, 303−307.

(48) Autschbach, J.; Le Guennic, B. *Chem.—Eur. J.* **2004**, *10*, 2581−2589.

(49) (a) Bühl, M.; Mauschick, F. T.; Terstegen, F.; Wrackmeyer, B. *Angew. Chem., Int. Ed.* **2002**, *41*, 2312−2315. (b) Bühl, M.; Mauschick, F. T. *Phys. Chem. Chem. Phys.* **2002**, *4*, 5508−5514. (c) Grigoleit, S.; Bühl, M. *Chem.—Eur. J.* **2004**, *10*, 5541−5552. (b) Bühl, M.; Schurhammer, R.; Imhof, P. *J. Am. Chem. Soc.* **2004**, *126*, 3310−3320.

(50) Laaksonen, A.; Wasylishen, R. E. *J. Am. Chem. Soc.* **1995**, *117*, 392−400.

(51) Malli, G.; Froese, C. *Int. J. Quantum Chem. Symp.* **1967**, *1*, 95−98.

(52) McFarlane,W. *J. Chem. Soc. A* **1968**, 1630−1634.

(53) Mitchell, T. N.; Amamria, A.; Fabisch, B.; Kuivila, H. G.; Karol, T. J.; Swami, K. *J. Organomet. Chem.* **1983**, *259*, 157−164.

(54) Birchall, T.; Pereira, A. *J. Chem. Soc., Chem. Commun.* **1972**, 1150−1151.

(55) Webster, J. R.; Jolly, W. L. *Inorg. Chem.* **1971**, *10*, 877−879.

(56) Lambert, J. B.; Zhao, Y.; Wu, H. W.; Tse, W. C.; Kuhlmann, B. *J. Am. Chem. Soc.* **1999**, *121*, 5001−5008.

(57) van den Berghe, E. V.; van der Kelen, G. P. *J. Organomet. Chem.* **1971**, *26*, 207−213.

(58) Lassigne, C. R.; Wells, E. J. *Can. J. Chem.* **1977**, *55*, 927−931.

(59) Burke, J. J.; Lauterbur, P. C. *J. Am. Chem. Soc.* **1961**, *83*, 326−331.

(60) McFarlane, W.; Wood, R. J. *J. Organomet. Chem.* **1972**, *40*, C17−C20.

(61) Kennedy, J. D.; McFarlane, W.; Pyne, G. S.; Clarke, P. L.; Wardell, J. L. *J. Chem. Soc., Perkin Trans. 2* **1975**, 1234−1239.

(62) Bagno, A.; Rastrelli, F.; Saielli, G. *J. Phys. Chem. A* **2003**, *107*, 9964−9973.

(63) Reed, C. A. *Acc. Chem. Res.* **1998**, *31*, 325−332.

(64) Kim, K. C.; Reed, C. A.; Elliott, D. W.; Müller, L. J.; Tham, F.; Lin, L. J.; Lambert, J. B. *Science* **2002**, *297*, 825−827.

(65) Lambert, J. B.; Lin, L.; Keinan, S.; Müller, T. *J. Am. Chem. Soc.* **2003**, *125*, 6022−6023.

(66) Cremer, D.; Olsson, L.; Reichel, F.; Kraka, E. *Isr. J. Chem.* **1993**, *33*, 369−385.

(67) McFarlane,W. *J. Chem. Soc. A* **1967**, 528−530.

(68) Flitcroft, N.; Kaesz, H. D. *J. Am. Chem. Soc.* **1963**, *85*, 1377−1380.

(69) Schumann, C.; Dreeskamp, H. *J. Magn. Reson.* **1970** *3*, 204−217.

(70) Petrosyan, V. S.; Pernin, A. B.; Reutov, O. A.; Roberts J. D. *J. Magn. Reson.* **1980**, *40*, 511−518.

(71) Petrosyan, V. S. *Prog. NMR Spectrosc.* **1977**, *11*, 115−148.

(72) Kuivila, H. G.; Kennedy, J. D.; Tien, R. Y.; Tyminski, I. J.; Pelczar, F. L.; Khan, O. R. *J. Org. Chem.* **1971**, *36*, 2083−2088.

(73) (a) Handy, N. C.; Cohen, A. J. *Mol. Phys.* **2001**, *99*, 403−412. (b) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868. (c) Becke, A. D. *Phys. Rev. A* **1988,** *38*, 3098−3100. (d) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785−789.

(74) (a) Mitchell, T. N. *J. Organomet. Chem.* **1974**, *70*, C1−C2. (b) Harris, R. K.; Mitchell, T. N.; Nesbitt, G. J. *Magn. Reson. Chem.* **1985**, *23*, 1080−1081.

CT050173K

JCTC Journal of Chemical Theory and Computation

The IMOMM (Integrated Molecular Orbitals/Molecular Mechanics) Approach for Ligand-Stabilized Metal Clusters. Comparison to Full Density Functional Calculations for the Model Thiolate Cluster $Cu_{13}(SCH_2CH_3)_8$

Alexander Genest,[†] André Woiterski,[†] Sven Krüger,[†] Aleksey M. Shor,[‡] and Notker Rösch*,[†]

Department Chemie, Theoretische Chemie, Technische Universität München, 85747 Garching, Germany, and Institute of Chemistry and Chemical Technology, Russian Academy of Sciences, 660049 Krasnoyarsk, Russian Federation

Received August 12, 2005

Abstract: To validate the IMOMM (integrated molecular orbitals/molecular mechanics) method for ligand-stabilized transition metal clusters, we compare results of this combined quantum mechanical and molecular mechanical (QM/MM) approach, as implemented in the program ParaGauss (Kerdcharoen, T.; Birkenheuer, U.; Krüger, S.; Woiterski, A.; Rösch, N. *Theor. Chem. Acc.* **2003**, *109*, 285), to a full density functional (DF) treatment. For this purpose, we have chosen a model copper ethylthiolate cluster, $Cu_{13}(SCH_2CH_3)_8$ in D_{4h} symmetry. The evaluation is based on 16 conformers of the cluster which exhibit single and bridging coordination of the ligands at the Cu_{13} cluster as well as various ligand orientations. For corresponding isomers, we obtained moderate deviations between QM and QM/MM results: 0.01−0.06 Å for pertinent bond lengths and up to ∼15° for bond angles. Ligand binding energies of the two approaches deviated less than 6 kcal/mol. The largest discrepancies between full DF and IMOMM results were found for isomers exhibiting short Cu−H and H−H contacts. We traced this back to the localization of different minima, reflecting the unequal performance of the DF and the force-field methods for nonbonding interactions. Thus, QM/MM results can be considered as more reliable because of the well-known limitations of standard exchange-correlation functionals for the description of nonbonding interactions for this class of systems.

Introduction

Accurate quantum chemical methods are restricted to calculations on small to mid-size systems. Large molecular species, like complexes with bulky ligands occurring in homogeneous catalysis or biomolecules, still have to be treated either by a less accurate approach or by a combination of quantum mechanical and molecular mechanical methods (QM/MM).[1−3] Combined approaches such as the integrated molecular orbitals/molecular mechanics (IMOMM) method,[4] a QM/MM variant, have found widespread use for treating systems where only a small part has to be described with high accuracy and the remaining part of the system can be considered as an "environment", exerting steric constraints or acting as a support. Typical examples of such systems are homogeneous catalysts with bulky ligands,[5,6] metal centers of heterogeneous catalysts at oxide surfaces or in zeolites cavities,[7] self-assembled monolayers at gold surfaces,[8] solvated complexes,[9−11] and large molecules of biological interest.[12−14]

Ligand-stabilized transition metal clusters can be viewed in analogy to metal complexes. The metal core features a

* Corresponding author tel. +49 89 289 13620; e-mail: roesch@ch.tum.de.

[†] Technische Universität München.

[‡] Russian Academy of Sciences.

10.1021/ct050202r CCC: $33.50
Published on Web 11/30/2005

rather complex electronic structure which, at least for metal centers in direct contact with ligands, easily responds to ligand binding. Thus, the electronic structure of the subsystem, which comprises the metal cluster proper and the metal−ligand bonds, has to be treated at a sophisticated level. In contrast, interligand interactions most often are dominated by van der Waals or electrostatic forces, which are amenable to modeling by a force-field approach.

Extending our previous work on small model complexes,[15] we apply here a recently developed implementation of the IMOMM approach to ligand-stabilized transition metal clusters. To the best of our knowledge, this is the first application of a QM/MM method to metal cluster compounds. To assess the accuracy of the IMOMM approach, we compare its results to those of the corresponding all-electron treatment. For this purpose, we have chosen the copper thiolate cluster $Cu_{13}(SCH_2CH_3)_8$ as a simple yet realistic example. In this way, we will validate our implementation and examine the performance of the IMOMM approach for a new class of systems. The choice of our model compound has been inspired by gold thiolate clusters, which recently attracted considerable interest as versatile building blocks of nanostructured materials and as realizations of quantum dots.[16,17] Although most of the experimental work on transition metal thiolate clusters is devoted to gold species, the corresponding copper compounds have also been synthesized.[18,19] The special interest in thiolate-stablized metal clusters is due to their rather simple synthesis, yielding stable products that are easy to handle, as well as the versatile chemistry of the thiolate ligands, which allows tailoring of the cluster surface for various purposes.[20,21]

This work is organized as follows. We briefly review the IMOMM method, proceed to describe specific features of the IMOMM implementation of the parallel density functional (DF) program ParaGauss,[22,23] and discuss other computational details. Then, we present the model cluster $Cu_{13}(SCH_2CH_3)_8$ and discuss its properties on the basis of DF calculations. Subsequently, we compare these results to those of IMOMM calculations which combine DF and force-field methods.

The IMOMM Implementation of ParaGauss

The QM/MM approach used in the present work is an adaptation of the IMOMM method,[4] which, besides standard IMOMM calculations, also allows one to treat ligated metal clusters.[15] In a QM/MM approach, one starts with partitioning a complex system *XY* into a "central" part *X*, treated by an accurate QM method, and its "environment" *Y*, described in an approximate way at a MM level. Correspondingly, one separates the total energy as

$$E(XY) = E_{QM}(X) + E_{MM}(Y) + E_{INT}(X,Y) \quad (1)$$

In the IMOMM method,[4] one approximates the interaction energy $E_{INT}(X,Y)$ between the two subsystems by its value at the lower level of accuracy (MM)

$$E_{INT}(X,Y) \approx E_{INT}^{appr}(X,Y) = E_{MM}(XY) - E_{MM}(X) - E_{MM}(Y) \quad (2)$$

This results in an "extrapolation" or "difference" scheme:[3,4]

$$E(XY) \approx E_{appr}(XY) = E_{QM}(X) + E_{MM}(XY) - E_{MM}(X) \quad (3)$$

For ligand-stabilized metal clusters, one has to cut covalent bonds (frontier bonds) of the ligands when one partitions the system *XY*. It is customary to cap the resulting "dangling" bonds of the QM region by "link atoms".[24] Different from the original approach,[4] the IMOMM implementation of ParaGauss[15] constrains the location $\vec{R}_2$ of link atoms to lie in the direction of the corresponding frontier bond from an atom at $\vec{R}_1$ (QM side) to an atom at $\vec{R}_3$ (MM side), by applying a fixed scaling factor *g*:[25]

$$\vec{R}_2 = \vec{R}_1 + g(\vec{R}_3 - \vec{R}_1) \quad (4)$$

Alternatively, one may keep these link bonds at a fixed length. Both procedures yield very similar results if the various parameters are suitably chosen.[15]

The IMOMM variant just described has been implemented in the parallel DF program package ParaGauss.[22,23] The implementation relies on the newly developed MM module MOLMECH[26] of ParaGauss and the geometry-optimizing module OPTIMIZER.[27] This new module of ParaGauss simplifies QM/MM calculations compared to the previous implementation,[15] which invoked an external MM program. QM/MM calculations carried out with MOLMECH benefit from the efficient symmetry treatment of ParaGauss.[28] The capability for QM/MM calculations is implemented in ParaGauss as an interface module which exchanges data between QM and MM modules on one hand and the OPTIMIZER module on the other. Relevant tasks are the preparation and distribution of data derived from a master input, the gathering of QM and MM contributions to energy gradients, and finally the calculation of the total QM/MM energy of the entire system *XY*.[15]

The module MOLMECH was designed to perform energy minimizations of molecules as well as of systems with two- and three-dimensional periodic boundary conditions for which atomic positions as well as unit cell parameters can be optimized. MOLMECH features a general open structure of force-field terms, which allows easy extension by new terms or new parameter sets. Electrostatic interactions of isolated molecules are treated either by a direct sum over atomic charges or by bond-centered dipoles as realized in the MM3 force field.[29] Electrostatic and van der Waals interactions of isolated systems are evaluated without cutoffs. Long-range electrostatic interactions in periodic systems, for example, in two- or three-dimensional arrays of ligated metal clusters, are calculated by Ewald techniques.[30,31] As this treatment of electrostatics is the computationally most-demanding part of a force field (FF) calculation, it has been parallelized employing the communication interface of ParaGauss.[22,23]

Computational Details

All QM calculations were carried out with the linear combination of Gaussian-type orbitals fitting-functions DF method[32] (LCGTO-FF-DF) as implemented in the parallel quantum chemistry package ParaGauss.[22,23] The geometry of the various systems was optimized using the local density

approximation (LDA)[33] for the exchange-correlation potential. LDA is well-known to yield reliable equilibrium geometries for transition metal compounds.[34−36] In contrast, LDA functionals tend to overestimate binding energies.[34] Therefore, we calculated energetic properties using the gradient-corrected BP86[37,38] functional (GGA = generalized gradient approximation) in a self-consistent single-point fashion.[39] We calculated the binding energy E_b per thionyl ligand SCH_2CH_3 as the difference of total energies:

$$E_b = E_{tot}[SCH_2CH_3] + \{E_{tot}[Cu_{13}] - E_{tot}[Cu_{13}(SCH_2CH_3)_8]\}/8 \quad (5)$$

The energy of the ethylthionyl ligand was determined in the conformation it featured in the cluster. For instance, the energy of SCH_2CH_3 in the eclipsed configuration was taken as reference for a copper cluster with eclipsed ligands (see below).

For five isomers (buis, buie, tuis, tuie, buos; see below for definitions), we probed the basis set superposition error (BSSE). We compared results for the ligand shell $(SCH_2CH_3)_8$ (with eight unpaired electrons) obtained without and with accounting for the Cu_{13} basis set. Correspondingly, we compared results for Cu_{13} without and with accounting for the basis set of the ligand shell. The total energy of Cu_{13} was lowered by up to 11 kcal/mol due to the ligand basis set, and the ligand shell gained up to 4.5 kcal/mol due to the Cu_{13} basis set. As our discussion later on is based on relative values of E_{tot}, namely, differences to that energy for configuration buos (see below), we estimate the BSSE of these relative energies to, at most, 4.3 kcal/mol (E_b will be affected by, at most, 0.5 kcal/mol), based on the differences between the BSSE results for the isomers just mentioned and the result for isomer buos.

To represent the Kohn−Sham orbitals, we applied the following basis sets: C (9s5p1d) → [5s4p1d],[40] S (12s9p2d) → [6s5p2d],[41] H (6s1p) → [4s1p],[40] and Cu (15s11p6d) → [6s4p3d].[42,43] All contractions were of generalized form, based on LDA atomic eigenvectors. The auxiliary basis set utilized in the LCGTO-FF-DF method to represent the electron charge density for treating the Hartree part of the electron−electron interaction was constructed by scaling s and p exponents of the orbital basis sets using a standard procedure.[32] On each atom, five p- and five d-type "polarization" exponents were added, chosen as geometric series with factors 2.5, starting with 0.1 and 0.2 au, respectively. For the numerical integration of the exchange-correlation contributions, a superposition of atom-centered spherical grids[44] was chosen, using angular grids which are locally accurate up to angular momentum $L = 19$.[45]

For the MM calculations, we used the same force field as that in our previous work[15] where parameters suitable for modeling copper thiolates have been proposed and evaluated. For the metal−metal interaction, only the van der Waals interaction was parametrized because these interactions cancel in the IMOMM scheme.[15] For the organic components, like the alkyl chains, MM3 FF[29] parameters describing stretching, bending, and torsion potentials were adopted. Geometries were relaxed until all components of the Cartesian gradients were smaller than 10^{-5} au and also the update step length dropped below that same value.

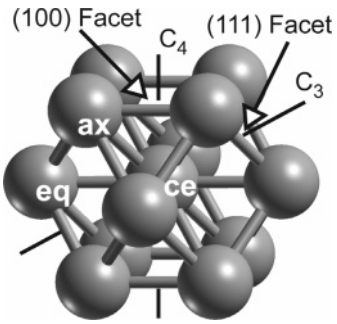


Figure 1. Cuboctahedral cluster Cu_{13} with labels of surface facets, pertinent C_n symmetry axes, and designators of the various atoms (ce = central, eq = equatorial, and ax = axial).

As expected, the QM/MM approach is computationally advantageous; for a given geometry, the time required for the electronic structure calculation (including the forces on the atoms) was reduced by about a factor of 2 compared to a QM calculation.

The Model Cluster $Cu_{13}(SCH_2CH_3)_8$

To examine the performance of the IMOMM approach for metal cluster compounds, we selected the copper thiolate cluster $Cu_{13}(SCH_2CH_3)_8$ in D_{4h} as a realistic but also computationally feasible model system. Ethylthionyl ligands are the shortest alkane thionyls for which steric interactions in the ligand shell are to be expected. In a recent study on $Au_{13}(SCH_3)_n$, we found that steric interactions are essentially absent between the smaller methylthionyl ligands.[46] For the Cu_{13} metal core, we adopted a cuboctahedral reference structure, comprising a central atom surrounded by a shell of 12 "surface" atoms. It is the smallest cluster that features a bulklike coordinated atom at its center. We preferred the cuboctahedral over the icosahedral shape of Cu_{13} because different ligand coordinations can easily be modeled. Also, the bare cluster Cu_{13} in I_h symmetry is 11 kcal/mol less stable; this result was not unexpected as the coinage metal cluster Au_{13} shows a similar preference.[39] During geometry optimization, we imposed D_{4h} symmetry constraints to restrict the structure of the ligand shell such that we were able to compare various ligand arrangements, optimized at both the DF and the QM/MM levels of theory. Overall, the cluster model chosen comprises all interactions present in larger transition metal thiolate clusters, yet it is simple enough to allow a full density functional treatment at the all-electron level for comparison.

Figure 1 introduces the designations of the various symmetry inequivalent Cu centers of the cluster. In D_{4h} symmetry, four surface atoms Cu_{eq} form a square in the horizontal (equatorial) symmetry plane, perpendicular to the C_4 main axis. Four metal atoms each form squares of (100) surface facets above and below that horizontal mirror plane; these centers Cu_{ax} are referred to as "axial". Finally, the central atom of the cluster is labeled as Cu_{ce}.

As reference, we optimized the bare metal core Cu_{13} both in O_h and D_{4h} symmetry. In O_h symmetry, all Cu−Cu bonds are equivalent and the LDA optimized bond length is 2.400

Å. The BP86 binding energy is 630 kcal/mol in total or 48.5 kcal/mol per atom. Because the highest occupied molecular orbital in O_h symmetry is only partially filled (t_{2g}^5), a Jahn−Teller distortion is expected, concomitant with a symmetry lowering. Applying D_{4h} symmetry constraints yielded two different isomers. The "round" isomer is bound with 631 kcal/mol and exhibits bonds that deviate, at most, 0.04 Å from those of the O_h reference. The other D_{4h} isomer features an overall oblate distortion where the equatorial atoms move outward ($Cu_{ce}-Cu_{eq} = 2.842$ Å) and the axial atoms move inward ($Cu_{ce}-Cu_{ax} = 2.296$ Å). With a BP86 atomization energy of 622 kcal/mol, this oblate structure is slightly less favorable than the O_h reference.

We adopted two starting configurations for the optimization of the ligated cluster $Cu_{13}(SCH_2CH_3)_8$. In the first case, ligands were singly coordinated to the axial Cu atoms—"on top" in the terminology of surface science. Alternatively, the ligands were attached to pairs of axial Cu atoms in 2-fold (bridge) coordination (Figures 2 and 3). We did not separately consider 3-fold coordination on the eight (111) facets of Cu_{13} because, in D_{4h} symmetry, ligands can move from bridging to (ideal) 3-fold positions. For the two starting configurations of D_{4h} symmetry, the S−C−C backbones of the ethylthionyl ligands lie in vertical mirror planes, limiting the number of possible conformations and, thus, facilitating a direct comparison of QM/MM and full QM results.

To distinguish different conformations of the ligands, we employ a labeling scheme that reflects the orientation of the ligands attached to the top facets of the cluster (Figures 2 and 3). First, the coordination of the ligands is classified as top (t) or bridging (b), according to the *starting* configuration; this designation is independent of where the ligands end up after optimization. For a ligand anchored on the metal cluster above the equatorial plane, the angle $Cu_{ax}-S-C$ (<180°) can be chosen to open upward (u) or downward (d) with respect to the C_4 main symmetry axis. In addition, the S−C−C moiety can be oriented toward (inward = i) or away from (outward = o) the C_4 axis. The last conformational degree of freedom in D_{4h} symmetry is the orientation of the terminal methyl group. It can be staggered (s) or eclipsed (e) with regard to the SCH_2 moiety. For example, the concatenated symbol "buos" designates a cluster isomer with bridging ligands (b), upward orientation (u) of the angle Cu−S−C, outward (o) opening of the angle S−C−C, and staggered conformation (s) of the methyl group. In summary, eight different conformers result for a given *initial* coordination mode (t or b), yielding a total of 16 conformers to be inspected.

According to experience with smaller compounds,[15] the C−C ligand bond has been chosen as the boundary between QM and MM regions. Thus, in the hybrid approach, the QM model was reduced to $Cu_{13}(SCH_3)_8$ and the terminal methyl groups of the ligands were treated at the MM level. The dangling C−C bonds were saturated by capping H atoms, using a constant ratio of the bond lengths, eq 4, with the scaling factor set to 0.709.[25] The boundary chosen between QM and MM partitions also accounts for the fact that charge transfer between these two regions is not included in the IMOMM model applied.

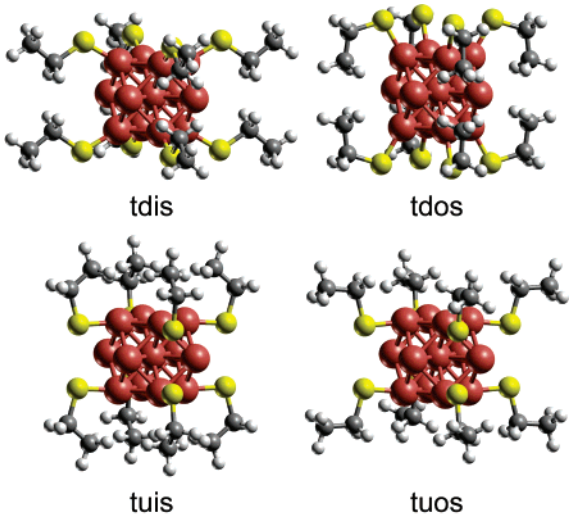


Figure 2. Four isomers of $Cu_{13}(SCH_2CH_3)_8$ with top coordination of the SCH_2CH_3 ligands and staggered orientation of the methyl group: ligands oriented downward−inward (tdis), downward−outward (tdos), upward−inward (tuis), and upward−outward (tuos).

QM Calculations on the Cluster $Cu_{13}(SCH_2CH_3)_8$

To set the stage for the evaluation of the QM/MM results on metal clusters, we first carried out QM reference calculations at the all-electron LDA level, optimizing structures for all 16 isomers. As visual inspection of optimized cluster geometries (Figures 2 and 3) reveals, ethylthionyl ligands are large enough so that steric effects play a role in the structure of the ligand shell of eight ligands assembled on the Cu_{13} cluster. Methyl end groups of top-coordinated ligands remain further from each other above the (100) facets because the ligands do not approach the cluster surface as closely as bridging ligands do. For the latter, this crowding effect is strongest for bdo isomers where the methyl end groups come close to each other near the horizontal symmetry plane (with shortest H−H contacts at 1.81 Å); there, methyl groups also get in close contact with Cu_{eq} atoms (with Cu−H = 1.97 Å). As top-coordinated ligands are further from the cluster surface, these contacts are weakened (shortest H−H at 2.18 Å and Cu−H at 2.43 Å) in tdo isomers (Figure 2).

For a detailed discussion of these observations, we have collected pertinent structure parameters in Table 1. We will first address "top" and then "bridge" isomers.

The cluster−ligand bond length $Cu_{ax}-S$ of top-coordinated ligands varies only slightly, between 2.10 and 2.13 Å, where the longer bonds are obtained for tdo isomers. Next-nearest copper−sulfur distances $Cu_{eq}-S$ exceed 3.7 Å for top ligands oriented downward, but they decrease to 3.1−3.2 Å for the upward orientation because ligands are shifted toward the equatorial plane (Figure 2). Ligand orientation also significantly affects the overall shape of the cluster as shown by the various Cu−Cu nearest-neighbor distances. The $Cu_{ce}-Cu_{eq}$ distance decreases along the series tui > tuo > tdi > tdo from ∼2.48 to 2.35 Å (Table 1). Concomitantly, the $Cu_{ce}-Cu_{ax}$ distance elongates, from ∼2.38 to ∼2.45 Å. Thus, the shape of the metal cluster core changes from oblate

Table 1. Characteristic Bond Lengths (in Å) of 16 Isomers of $Cu_{13}(SCH_2CH_3)_8$ Optimized with D_{4h} Symmetry Constraints[a]

isomer	$Cu_{ce}-Cu_{eq}$	$Cu_{ce}-Cu_{ax}$	$Cu_{eq}-S$	$Cu_{ax}-S$	$S-C_1$	C_1-C_2	$Cu_{ax}-S-C_1$
buis	2.339	2.662	2.298	2.193	1.853	1.500	118.4
buie	2.357	2.628	2.321	2.188	1.860	1.514	117.7
buos	2.345	2.604	2.365	2.201	1.860	1.503	115.4
buoe	2.356	2.614	2.326	2.196	1.863	1.516	119.4
bdis	2.283	2.573	2.837	2.236	1.847	1.510	109.5
bdie	2.286	2.572	2.835	2.234	1.850	1.523	109.8
bdos	2.299	2.603	3.281	2.254	1.815	1.493	113.6
bdoe	2.268	2.615	3.212	2.249	1.822	1.512	118.0
tuis	2.471	2.383	3.136	2.119	1.826	1.502	102.8
tuie	2.485	2.366	3.203	2.118	1.834	1.527	97.7
tuos	2.451	2.394	3.088	2.105	1.838	1.509	101.0
tuoe	2.449	2.400	3.077	2.103	1.835	1.525	104.8
tdis	2.402	2.409	3.708	2.101	1.823	1.510	116.3
tdie	2.403	2.408	3.707	2.101	1.824	1.522	116.6
tdos	2.345	2.459	4.113	2.134	1.817	1.501	104.0
tdoe	2.351	2.446	4.092	2.127	1.819	1.520	108.4
bdos[b]	2.859	2.500	2.173	2.244	1.850	1.503	138.6
bdoe[b]	2.868	2.500	2.168	2.239	1.859	1.512	138.7
tuis	2.471	2.383	3.136	2.119	1.826	1.502	102.8
tuie[b]	2.481	2.371	3.158	2.112	1.826	1.526	101.7

[a] For the designation of the various atoms, see Figure 1; for the designation of the isomers, see the text. [b] New isomers found with the help of IMOMM results.

to prolate. Clusters with "upward" oriented ligands exhibit longer distances from the center to the equatorial Cu atoms than in the corresponding conformation with the ligands oriented "downward". For instance, in td conformers, the $Cu_{ce}-Cu_{eq}$ bonds are ~0.1 Å longer than for the corresponding tu isomers (cf. tdi vs tui). Correspondingly, the $Cu_{ce}-Cu_{ax}$ distance is ~0.05 Å longer in td isomers than in the corresponding tu conformers.

The orientation of the methyl group, staggered or eclipsed, affects bond distances in a minor way, typically by 0.01 Å or less, but in some cases, these bonds can differ by up to ~0.02 Å between two corresponding isomers. Bond lengths within the ligands vary in similarly narrow ranges: 1.82−1.84 Å for $S-C_1$ and 1.50−1.53 Å for C_1-C_2. Bond angles for $Cu_{ax}-S-C_1$ are ~100° for tu, ~105° for tdo, and ~116° for tdi isomers.

The equilibrium geometries of clusters, where the structure optimization started with bridge-coordinated ligands, show rather different trends. An inspection of Figure 3 as well as a comparison of the distances $Cu_{ax}-S$ and $Cu_{eq}-S$ reveals that true 2-fold coordination is obtained only for the four types of isomers with downward-oriented ligands (bd). For these isomers, the $Cu_{ax}-S$ bond length is 2.23−2.25 Å, while the $Cu_{eq}-S$ distance remains considerably longer, ~2.8 Å for bdi isomers and ~3.2 Å for bdo structures (Table 1). Bridge-hollow coordination is found for the four types of bu isomers. There, the S atom lies still closer to the axial copper atoms, with $Cu_{ax}-S$ bonds of 2.19−2.20 Å, but the $Cu_{eq}-S$ contacts are only 0.11−0.17 Å longer, giving rise to some bonding interaction in these bridge-hollow coordination modes. In agreement with steric considerations, S atoms shift furthest to the 3-fold coordination site for bui rotamers (Figure 3). The ligands try to avoid the steric stress above the (100) facet by moving the S atom closer to the Cu_{eq} centers. $Cu_{ax}-S$ bonds of bridging ligands are systematically

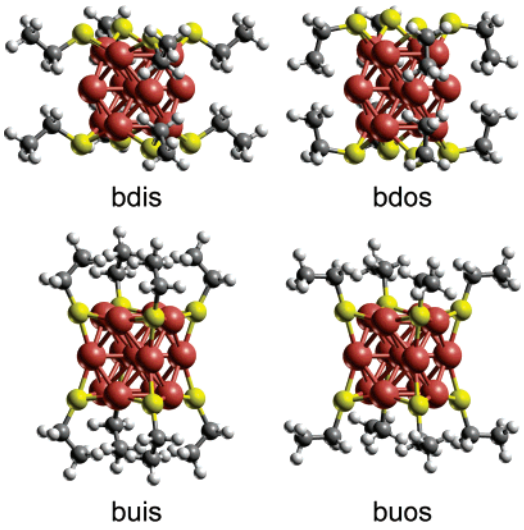


Figure 3. Four isomers of $Cu_{13}(SCH_2CH_3)_8$ with bridge coordination of the SCH_2CH_3 ligands and staggered orientation of the methyl group: ligands oriented downward−inward (bdis), downward−outward (bdos), upward−inward (buis), and upward−outward (buos).

longer than those of top-coordinated ligands by ~0.1 Å, as previously found for $Au_{13}(SMe)_n$ clusters.[46] A comparison of $Cu_{ce}-Cu_{eq}$ and $Cu_{ce}-Cu_{ax}$ bonds of a given isomer shows that all these structures are prolate, with the $Cu_{ce}-Cu_{ax}$ bonds longer by 0.26−0.35 Å; however, there is no clear trend as was found in the clusters with top-coordinated ligands. $Cu_{ce}-Cu_{eq}$ distances are 2.34−2.36 Å for bu isomers (with upward-oriented bridging ligands) and 2.27−2.30 Å for bd isomers (with downward-oriented ligands). The orientation of the methyl group affects Cu−Cu bonds in a similarly minor fashion as that in the top-coordinated clusters, with two exceptions: the two pairs buis−buie and bdos−bdoe feature changes of the $Cu_{ce}-Cu_{eq}$ and $Cu_{ce}-Cu_{ax}$ bonds, up

Table 2. Total Energies E_{tot} Relative to That of the Isomer buos and Binding Energies E_b Per Ligand for the Various Isomers of $Cu_{13}(SCH_2CH_3)_8$ as well as Corresponding Energy Differences ΔE_{tot} and ΔE_b between Staggered (s) and Eclipsed (e) Conformers of a Given Ligand Coordination Mode[a]

isomer	E_{tot}	ΔE_{tot}	E_b	ΔE_b
buis	1.7	−25.3	62.9	0.6
buie	26.9		62.3	
buos	0.0	−23.3	63.1	0.3
buoe	23.3		62.8	
bdis	30.1	−25.3	59.4	0.6
bdie	55.4		58.8	
bdos	64.7	−21.3	55.0	0.1
bdoe	86.0		55.0	
tuis	106.2	−40.4	49.8	2.5
tuie	146.6		47.4	
tuos	102.8	−25.3	50.3	0.6
tuoe	128.1		49.7	
tdis	122.7	−25.8	47.8	0.6
tdie	148.5		47.1	
tdos	151.1	−18.9	44.2	−0.2
tdoe	170.0		44.4	
bdos[b]	2.2	−23.5	62.8	0.3
bdoe[b]	25.7		62.5	
tuis	106.2	−31.3	49.8	1.3
tuie[b]	137.4		48.5	

[a] Energies in kcal/mol. [b] New isomers found with the help of IMOMM results.

to 0.03 Å, in opposite directions. The bond lengths within the ligands also vary within similar margins as those for cluster isomers with top coordination: 1.82−1.86 Å for $S-C_1$ and 1.49−1.52 Å for C_1-C_2. $Cu_{ax}-S-C_1$ angles are 114−120° for bu as well as bdo isomers and are 110° for bdi isomers.

Comparing clusters of the same ligand orientation, one finds $Cu_{ce}-Cu_{eq}$ bonds for top coordination longer than for bridge coordination and $Cu_{ce}-Cu_{ax}$ bonds shorter. $Cu_{ce}-Cu_{eq}$ bonds can differ by 0.08−0.13 Å, and $Cu_{ce}-Cu_{ax}$ bonds can differ by 0.14−0.28 Å (Table 1).

In Table 2, we compare various energetic parameters of the 16 isomers. The total energy is referenced to that of the most stable isomer, buos. In general, isomers with bridging ligands are 90−120 kcal/mol more stable than isomers with top-coordinated ligands of the same orientation. While the total energies of isomers with bridging ligands span an interval of ∼85 kcal/mol, the energies of isomers with top-coordinated ligands scatter over an interval of ∼70 kcal/mol. The most stable isomer with top-coordinated ligands, tuos, is 17 kcal/mol less stable than the least stable isomer with bridge-coordinated ligands, bdoe, and more than 100 kcal/mol less stable than the most stable isomer, buos. These energy differences result from variations of the ligand binding energies E_b, which are 11−15 kcal/mol for a given ligand orientation (Table 2).

Among the "top" conformers, tdo isomers exhibit the smallest ligand binding energies, ∼44 kcal/mol, followed by tdi isomers with ∼47 kcal/mol; tu isomers have the largest ligand binding energies, ∼50 kcal/mol. The binding energy of bridge-coordinated ligands varies typically between 55 and 63 kcal/mol and shows the same ordering as that for top-coordinated ligands. bdo isomers have the lowest binding energy, 55 kcal/mol, followed by bdi isomers with 59 kcal/mol and bu isomers with ∼63 kcal/mol. Because E_b values are referenced to corresponding rotamer structures, staggered and eclipsed, they show more clearly than E_{tot} values the direct ligand−cluster interaction including steric effects. This conclusion is supported by the very small differences ΔE_b between corresponding values of staggered and eclipsed structures (Table 2).

As expected, the eclipsed form of the ethyl end group leads to a higher total energy. For isolated ethylthionyl ligands SCH_2CH_3, the energy difference between staggered and eclipsed conformations was calculated at 2.6 kcal/mol. The corresponding rotational barrier of ethylthiol $HSCH_2CH_3$ is calculated slightly higher, at 3 kcal/mol. For a cluster with eight ligands, these values extrapolate to ∼21 and 24 kcal/mol, respectively. Accordingly, the energies of most pairs of rotamers between staggered (s) and eclipsed (e) conformations differ by 23−25 kcal/mol (see ΔE_{tot}, Table 2). For the bdo and tdo isomers, this energy difference ΔE_{tot} is ∼3 and 5 kcal/mol, respectively, smaller (by absolute value) than the average value of 24 kcal/mol; however, these energy variation values translate into binding energy changes of less than 1 kcal/mol per ligand (see ΔE_b, Table 2). Only the tui rotamers are separated notably further in energy as the eclipsed rotamer is 40 kcal/mol less stable than the corresponding staggered structure. This increase of the energy difference is due to specific ligand−ligand interactions which are enforced by the constraints of the ligand conformation (Figure 2). Indeed, in the tuie conformation, the energy of the ligand shell $(SCH_2CH_3)_8$ (in a configuration with eight unpaired spins) is 13.6 kcal/mol *destabilized* relative to the energy of eight isolated ligands in the eclipsed conformation. Yet, a single ethylthionyl in the optimized conformation of the tuie isomer is only 0.1 kcal/mol less stable than the free ligand. In contrast, the ligand shell $(SCH_2CH_3)_8$ is *stabilized* by 5.1 kcal/mol relative to the energy of eight isolated ligands in a staggered conformation. Thus, the unusual high destabilization of the eclipsed tui rotamer derives from an unfavorable interligand interaction.

As expected, we determined a doublet ground state for all but two isomers of the cluster $Cu_{13}(SCH_2CH_3)_8$, reflecting the odd number of electrons of the system. The tdo pair of isomers was found to be more stable in the quartet state. For structures with bridging ligands, the gap between highest occupied and lowest unoccupied spin−orbitals is 0.5−0.9 eV, whereas it is considerably smaller, 0.05−0.2 eV, for isomers with terminal ligand coordination, rationalizing to some extent the exceptional quartet state of two t-type structures.

Finally, we mention additional cluster isomers of the td type with an overall oblate shape (not listed in the tables), which we found when we tried to use the flat, bare Cu_{13} as an underlying cluster core. Because these cluster conformations represent states of higher energy compared to their prolate congeners (E_{tot} 3.7 kcal/mol higher for tdie and tdoe,

Table 3. Characteristic Geometric Parameters of the Cu_{13} Cluster Core and the Cluster−Ligand Interface of 16 Conformers of $Cu_{13}(SCH_2CH_3)_8$ from QM/MM Calculations and Deviations δ(QM/MM-QM) from the Corresponding QM Results[a]

isomer	$Cu_{ce}-Cu_{eq}$	δ	$Cu_{ce}-Cu_{ax}$	δ	$Cu_{eq}-S$	δ	$Cu_{ax}-S$	δ	$Cu_{ax}-S-C_1$	δ
buis	2.368	0.029	2.644	−0.018	2.287	−0.011	2.187	−0.006	124.1	5.6
buie	2.370	0.013	2.631	0.003	2.314	−0.006	2.188	0.000	121.9	4.2
buos	2.357	0.012	2.619	0.015	2.322	−0.043	2.193	−0.008	118.4	3.1
buoe	2.358	0.003	2.620	0.006	2.320	−0.006	2.192	−0.004	118.4	−1.0
bdis	2.276	−0.007	2.578	0.005	2.874	0.036	2.234	−0.002	107.2	−2.3
bdie	2.276	−0.009	2.579	0.006	2.882	0.047	2.234	−0.001	107.0	−2.9
bdos	2.807	0.508	2.509	−0.095	2.177	−1.104	2.240	−0.014	138.8	25.2
bdoe	2.811	0.543	2.507	−0.108	2.177	−1.035	2.235	−0.014	138.3	20.2
tuis	2.485	0.014	2.368	−0.015	3.163	0.027	2.111	−0.007	113.1	10.2
tuie	2.483	−0.002	2.369	0.003	3.169	−0.034	2.109	−0.008	116.7	18.9
tuos	2.482	0.030	2.376	−0.018	3.132	0.044	2.112	0.006	102.9	1.9
tuoe	2.482	0.033	2.376	−0.024	3.132	0.055	2.111	0.008	103.0	−1.9
tdis	2.394	−0.007	2.414	0.005	3.733	0.025	2.103	0.003	113.0	−3.3
tdie	2.395	−0.008	2.414	0.005	3.735	0.027	2.103	0.002	112.9	−3.8
tdos	2.349	0.005	2.442	−0.017	4.072	−0.041	2.116	−0.018	116.1	12.1
tdoe	2.339	−0.011	2.445	−0.001	4.069	−0.023	2.115	−0.013	121.8	13.3
bdos[b]	2.807	−0.053	2.509	0.009	2.177	0.005	2.240	−0.004	138.8	0.2
bdoe[b]	2.811	−0.057	2.507	0.007	2.177	0.008	2.235	−0.004	138.3	−0.4
tuis	2.485	0.014	2.368	−0.015	3.163	0.027	2.111	−0.007	113.1	10.2
tuie[b]	2.483	0.002	2.369	−0.002	3.169	0.011	2.109	−0.003	116.7	14.9

[a] Bond lengths and their differences in Å; angles and their differences in degrees. [b] Differences to new DF isomers found with the help of IMOMM results.

5.5 kcal/mol for tdis, and 7.4 kcal/mol for tdos), we will not discuss them further.

IMOMM Calculations on the Cluster $Cu_{13}(SCH_2CH_3)_8$

To assess the applicability and accuracy of the IMOMM QM/MM approach for metal cluster compounds, we will now compare IMOMM results to those obtained previously in all-electron QM calculations. At first, we used the very same initial structures for the QM/MM geometry optimizations as those used previously for the pure QM optimizations—to locate, as far as possible, the same local minima. We will begin with discussing the results of these optimizations. For some isomers, this strategy failed to identify analogous local minima, and we had to expand our search, as will be detailed later on.

To characterize the consequences of the different computational methods, we will start with a discussion of the largest structural deviations. The geometric parameters of IMOMM-optimized cluster compounds are displayed in Table 3, together with the deviations from the corresponding QM results. An inspection of the bond distances and their deviations reveals that the largest differences between QM and QM/MM results occur for the bdo isomers; below, we will discuss these structures separately. For the remaining structures, the average absolute deviations are 0.013 Å (0.033 Å) for $Cu_{ce}-Cu_{eq}$, 0.010 Å (0.024 Å) for $Cu_{ce}-Cu_{ax}$, 0.030 Å (0.055 Å) for $Cu_{eq}-S$, 0.006 Å (0.018 Å) for $Cu_{ax}-S$, and 6.0° (18.9°) for $Cu_{ax}-S-C_1$; the maximum absolute deviations are given in parentheses. From these values, one concludes that the IMOMM method works rather well for such ligated metal cluster compounds. Distances $Cu_{ax}-S$ are particularly well-reproduced. In contrast, discrepancies in the angles $Cu_{ax}-S-C_1$ indicate a propensity for easy deformation in this structural characteristic. As these angles are not stabilized by any additional direct bond, they react strongly on (small) changes in the environment and, therefore, can be used as sensitive indicators for (other) very small structural discrepancies. According to this criterion, the isomers bdo, tui, and tdo deserve special attention (Table 3) as the discrepancies in the angles range from 10 to 25°. Also, the bui isomers, with differences in the $Cu_{ax}-S-C_1$ angle of 4−6°, can be mentioned in this context.

We first turn to the bdo isomers, which show the largest structural differences between IMOMM and full QM calculations. In the QM/MM structures, the $Cu_{eq}-S$ distances are significantly elongated, more than 1 Å, compared to the corresponding QM structures (Table 3). This strong discrepancy reflects the displacement of the ligands from bridge sites in the QM-optimized structure to 3-fold hollow positions in the QM/MM case (Figure 4). Also, the structures of the Cu_{13} core differ noticeably between the two types of calculations. In the IMOMM calculations, the $Cu_{ce}-Cu_{eq}$ distances are ∼0.5 Å longer and the axial bonds $Cu_{ce}-Cu_{ax}$ are 0.1 Å shorter. Thus, the shape of the cluster core as determined by the QM/MM calculations is quite similar to that of the oblate bare cluster Cu_{13}. Recall that the latter isomer of the bare cluster is only ∼10 kcal/mol less stable than the prolate isomer. These very substantial differences for the bdo isomers between the results of the two computational methods, which obviously do not present the same minimum at the potential energy surface, can be traced to the corresponding ligand arrangements. In the all-electron case, the ligands of the bdo rotamers wrap around the cluster surface (Figure 4), resulting in rather short contacts (<2.5 Å) between the methyl groups of the ligands and the Cu_{eq} centers. In the bdoe conformer, the corresponding $Cu_{eq}-H$ contacts are just 1.97 Å, but this distance is 2.31 Å in the bdos isomer. One expects this

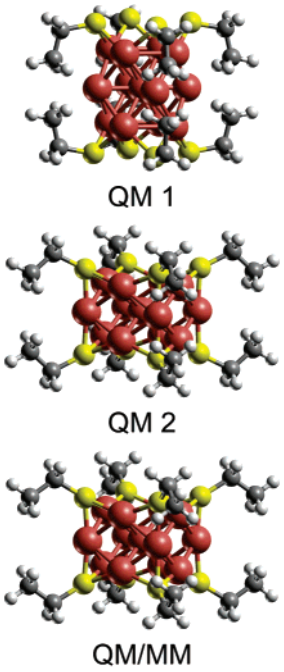


Figure 4. Geometries of the bdos isomer with bridge-coordinated ligands in "downward" orientation with outward-pointing methyl tail groups. Structures from full DF (QM1) and QM/MM calculations which have been started from the same initial structure as well as the DF (QM2) result which has been obtained when starting from the QM/MM geometry.

nonbonding interaction to be clearly repulsive. Indeed, the QM/MM structure (Figure 4) seems to imply a strong repulsion as the $Cu_{eq}-H$ contacts elongate substantially, to 3.29 Å in the bdoe conformer and to 3.57 Å in the bdos conformer.

These significant differences between the results for the bdo structures can be attributed to the different representation of nonbonding interactions by the DF and FF methods. The currently popular exchange-correlation approximations of LDA and GGA types are unable to account for dispersion interaction in a quantitative fashion,[47] although some success has recently been claimed with specially parametrized functionals as well as a time-dependent density-functional theory approach.[48−52] For the LDA functional applied here, this methodological deficiency becomes manifest in too-short nonbonding contacts[53] which may be rationalized by an artificial attraction due to density overlap;[54] this failure is also observed for GGA functionals.[54] In addition, a common feature of all mathematically local density functionals is the missing dispersion interaction.[47,54] On the other hand, force fields explicitly account for van der Waals interaction via their parametrization. Thus, for short nonbonding contacts, one can expect proper repulsion and, consequently, more reliable results from a FF (or a QM/MM) approach than from a pure DFT-based method. Furthermore, if nonbonding contacts are present, the risk of locating different minima in QM/MM and DF optimizations is increased as a result of artificial stabilization of nonbonding contacts in LDA or GGA calculations.[54]

As a check of this hypothesis, we evaluated the difference ΔE_{MM} of the force-field energy contributions at QM- and QM/MM-optimized geometries for the isomers bdo, buo, tui, and tdo:

$$\Delta E_{MM} = [E_{MM}(XY) - E_{MM}(X)]_{QM\ geometry} - [E_{MM}(XY) - E_{MM}(X)]_{QM/MM\ geometry} \quad (6)$$

We chose the isomers with the largest deviations between QM and QM/MM and one pair with small structure differences as counterexamples (buo). More than 95% of ΔE_{MM} derives from van der Waals interaction. Just as expected, ΔE_{MM} values for the bdo isomers are unusually large, more than 400 kcal/mol. Also, the ΔE_{MM} of isomer tuie is quite large, 330 kcal/mol. The analogous values for other conformers are notably smaller, 109 kcal/mol for tdoe, 52 kcal/mol for tdos, and 60 kcal/mol for tuis. As expected, ΔE_{MM} values are very small, 1.1 and 6.6 kcal/mol, for the buo counterexamples.

According to these findings, QM and QM/MM results for other isomers with short interligand H−H or Cu−H contacts should also differ. Indeed, bui, tui, and tdo conformers (in addition to the bdo structures) exhibit such relatively short interligand contacts. Cu−H contacts of the tdo isomers from the all-electron QM calculations (with Cu_{ax}: tdos − 2.77 Å and tdoe − 2.43 Å) are quite a bit longer than in the bdo isomers (2.31 and 1.97 Å; see above). Consequently, differences between QM and QM/MM results for bond lengths were considerably smaller for tdo than for bdo isomers, although $Cu_{ax}-S-C$ bond angles increased notably in the tdo isomers, by 12−13° (Table 3). In the same spirit, relatively short H−H contacts of about 2.3 and 2.5 Å for the tuis and tuie isomers, respectively, did not prevent good agreement among bond distances obtained from QM and QM/MM calculations but were reflected in larger values of $Cu_{ax}-S-C$ angles from QM/MM calculations, 10° (tuis) and 19° (tuie). For bui isomers, the structural trends seem comparable to those of the tui isomers, but the changes in the $Cu_{ax}-S-C$ angles are much smaller (Cu−H = 3.11 and 3.16 Å and H−H = 2.42 and 1.91 Å for buie and buis, respectively). Because bridge-coordinated ligands are closer to the "surface" of the cluster core, they have to bend less to form H−H contacts comparable to those of tui structures. For the tuie isomer, even a very short $Cu_{ax}-H$ contact of 2.085 Å was obtained in the full QM calculation.

The structures with close Cu−H and H−H contacts are the very same structures that we previously had singled out with the help of the criterion of the $Cu_{ax}-S-C_1$ angles. Whereas the $Cu_{ax}-S-C_1$ angles from QM/MM and full QM calculations differ noticeably in the bui, tdo, and tui conformers, bond distances deviate, at most, by 0.04 Å. Apparently, the ligands are just differently oriented, but the individual structures of both the ligands and the cluster core remain largely unchanged. Indeed, the shape of the clusters stays oblate for "bridge" coordination and prolate for "top" coordination. Also, the IMOMM approach yields the same 2- or 3-fold coordination as that of the full QM calculations.

The differences between the results from the QM and QM/MM approaches are reflected by the energetics as well (Table 4). Also, for the IMOMM calculations, we use the buos structure as an energy reference. Bridge-coordinated

Table 4. Relative Total Energies E_{tot}, Binding Energies E_b Per Ligand from QM/MM Calculations on 16 Conformers of $Cu_{13}(SCH_2CH_3)_8$, and Corresponding Energy Differences ΔE_{tot} and ΔE_b between Staggered (s) and Eclipsed (e) Conformers of a Given Ligand Coordination Mode. Also Shown Are the Differences δE_{tot} and δE_b from the Corresponding QM Energies[a]

isomer	E_{tot}	ΔE_{tot}	δE_{tot}	E_b	ΔE_b	δE_b
buis	−6.9	−30.9	8.6	66.7	−0.2	−3.8
buie	24.0		2.9	67.0		−4.6
buos	0.0	−34.1	0.0	65.8	0.2	−2.7
buoe	34.1		−10.9	65.7		−2.9
bdis	32.7	−34.7	−2.7	61.7	0.2	−2.4
bdie	67.4		−12.1	61.5		−2.7
bdos	−3.0	−39.7	67.7	66.2	0.8	−11.2
bdoe	36.6		49.3	65.4		−10.4
tuis	100.2	−45.8	6.0	53.3	1.6	−3.5
tuie	146.0		0.6	51.7		−4.3
tuos	100.7	−34.8	2.1	53.2	0.2	−3.0
tuoe	135.5		−7.4	53.0		−3.3
tdis	126.2	−34.6	−3.6	50.1	0.2	−2.3
tdie	160.8		−12.3	49.8		−2.7
tdos	127.8	−43.0	23.3	49.9	1.3	−5.6
tdoe	170.8		−0.8	48.6		−4.2
bdos[b]	−3.0	−39.7	5.2	66.2	0.8	−3.4
bdoe[b]	36.6		−10.9	65.4		−2.9
tuis	100.2	−45.8	6.0	53.3	1.6	−3.5
tuie[b]	146.0		−8.6	51.7		−3.2

[a] Energies in kcal/mol. [b] Differences to new DF isomers found with the help of IMOMM results.

clusters have IMOMM energies that are up to 68 kcal/mol higher than that of the reference, spanning an interval of 74 kcal/mol, compared to 86 kcal/mol at the full QM level. The relative total energies of top-coordinated clusters fall into the range from 100 to 171 kcal/mol, compared to 103−170 kcal/mol at the full QM level. The energy separation of clusters with bridge- and top-coordinated ligand shells is somewhat more pronounced at the QM/MM level (33 kcal/mol) than at the full QM level (16 kcal/mol).

The buos structure features a low energy also at the IMOMM level, but the buis and bdos structures are somewhat lower in energy. At the QM/MM level, the buis structure is 8.6 kcal/mol stabilized compared to the full QM calculation (δE_{tot}, Table 4). In fact, the total energies from both types of calculations correlate reasonably well (Figure 5), with differences δE_{tot} typically ranging from −12 to 9 kcal/mol, with three rather notable exceptions: the structures bdos (68 kcal/mol), bdoe (49 kcal/mol), and tdos (23 kcal/mol). Below, we will discuss these structures in more detail.

As in the full QM calculations, isomers with staggered methyl substituents are always more stable than those with an eclipsed orientation of the methyl groups. The differences ΔE_{tot} from the IMOMM calculations, ranging now from 31 to 46 kcal/mol, are notably larger than those from the full QM calculations. IMOMM values typically are close to 34 kcal/mol, 10 kcal/mol larger than typical QM values (Table 4). This difference is traced back to corresponding differences for the isolated thionyl ligands: In the QM/MM approach, the rotational barrier of the ethyl end group is calculated at 4.0 kcal/mol, whereas a full DF treatment yields a value of 2.6 kcal/mol. Test calculations on ethane yielded rotational barriers of 2.9 kcal/mol for QM/MM and 2.6 kcal/mol for QM, in good agreement with experimental values (2.9 kcal mol[55]).

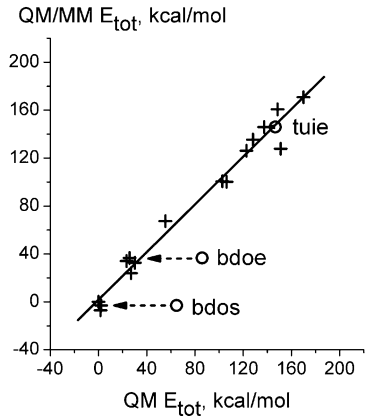


Figure 5. Correlation of E_{tot}, QM vs QM/MM values (in kcal/mol). Crosses denote results of corresponding minima; circles refer to deviating minima obtained when invoking the same starting condition in the optimization. For details, see text.

The ligand binding energies E_b from full QM calculations are very well reproduced with the IMOMM approach. Ligand binding energies of bridge-coordinated structures from IMOMM calculations, 62−67 kcal/mol, are again notably larger than those of top-coordinated structures, 49−53 kcal/mol. These E_b values show the same independence of the orientation of the methyl substituents (staggered vs eclipsed) when referenced to the appropriate rotamer structure; the largest difference ΔE_b is 1.6 kcal/mol. An inspection of Table 4 reveals that the ligand binding energies E_b from the IMOMM calculations are consistently larger than those from full QM calculations, by 2.3−5.6 kcal/mol if one leaves aside the bdo isomers with δE_b values of −11.2 and −10.4 kcal/mol, respectively. These differences are mainly due to approximating ethylthionyl ligands by their methylthionyl congeners in the QM subsystem of the IMOMM calculations.[15] Slight variations among the (regular) ligand binding energies, for example, from $\delta E_b = -2.7$ kcal/mol for buos to $\delta E_b = -3.8$ kcal/mol for buis, stabilize the latter structure and cause it to become the ground state of the QM/MM calculations.

We did not observe major differences in the electronic structure between QM and QM/MM results, although the QM system in the IMOMM approach is reduced to $Cu_{13}(SCH_3)_8$. As in the QM calculations, doublet ground states were obtained for all isomers except for the pair tdo, for which, again, quartet states were determined. Gaps between highest occupied and lowest unoccupied spin−orbitals amounted to 0.6−0.85 eV for structures with bridge-coordinated ligands and to 0.05−0.1 eV for structures with terminally coordinated ligands, again, in good agreement with the QM results. Thus, truncation of the ligand at the first CC bond also preserves essential features of the electronic structure of the cluster in the QM/MM calculation.

After this examination of the QM/MM results and their deviations from a full DF treatment, we note adequate overall similarity between the energetics at the two levels of theory. The bdo isomers with their artificially close Cu−H contacts at the all-electron DF level have already been identified as special cases when we analyzed the structures. Above, we had concluded that IMOMM structures are more realistic than the pure QM structures, which are significantly higher in energy for bdo isomers (by 50−70 kcal/mol; Table 4). Given these large energy differences, it seems worth studying whether the more realistic energetics at the QM/MM level in fact produce a local minimum that corresponds to the one located previously at the all-electron QM level.

To probe this conjecture, we started all-electron QM structure optimizations from *all* geometries optimized at the IMMOM level. Indeed, for 3 of the 16 isomers, we found new minima with lower full QM total energies than obtained previously. It is not too surprising that the two special structures bdos and bdoe were among them; the new QM structures turned out to be substantially more stable (by 62.5 and 60.3 kcal/mol, respectively) than the old QM structures. The third case was the isomer tuie, also discussed before as a case with unusually close nonbonding contacts (see above); its new QM structure is, by 9.2 kcal/mol, more stable. As expected, ΔE_{MM} values characterizing the van der Waals repulsion decreased substantially for the new QM structures (bdos: 11 kcal/mol; bdoe: 40 kcal/mol; tuie: 177 kcal/mol). The corresponding structure and energy data for the bdo and tui isomers are displayed in the lower sections of Tables 1−4. Note that there is only one QM structure for the tuis isomer; the corresponding data are shown to allow a full comparison of all table entries.

The new bdo structures are now the only bridge-coordinated systems for which we found an oblate cluster core (Figure 4, QM2). Their ligands are 3-fold bound, as in the bu isomers. The S−C bond lengths of the new structures now fit better with values for other clusters with bridging ligands. The new, full QM structures of the bdo isomers agree well with the corresponding QM/MM structures, with the largest differences (0.053−0.057 Å) occurring for the $Cu_{ce}-Cu_{ax}$ distances (Table 3). The new QM structure of the tuie isomer is structurally quite similar to the old one, also yielding a relatively large difference of 14.9° to the QM/MM result of the sensitive angle $Cu_{ax}-S-C_1$ (Table 3). Now, the average absolute (and maximum) deviations between the QM and QM/MM results of *all* isomers are 0.018 Å (0.057 Å) for $Cu_{ce}-Cu_{eq}$, 0.010 Å (0.024 Å) for $Cu_{ce}-Cu_{ax}$, 0.026 Å (0.055 Å) for $Cu_{eq}-S$, 0.006 Å (0.018 Å) for $Cu_{ax}-S$, and 5.1° (14.9°) for $Cu_{ax}-S-C_1$.

The three new isomers also fit the energetic characteristics of their congeners very well. The energy difference ΔE_{tot} between the staggered and eclipsed tui conformers is now reduced to −31.3 kcal/mol from, previously, −40.4 kcal/mol (Table 2). Most noticeable is the agreement of the new QM energies for the bdo isomers with the corresponding QM/MM values; the δE_{tot} values (bdos: 5.2 kcal/mol; bdoe: −10.9 kcal/mol) now fall in the normal range (see above and Table 4). With these new total energies, the correlation between QM and QM/MM improves drastically; the regression coefficient r^2 for the values of E_{tot} increases from 0.87 for the original data to, now, 0.98 (Figure 5). A similar improvement is observed for the differences δE_b between QM and QM/MM ligand binding energies. The deviation, previously ∼10 kcal/mol, decreased to less than 4 kcal/mol (bdos: −3.4 kcal/mol; bdoe: −2.9 kcal/mol; Table 4).

From this reoptimization of QM structures, we conclude that the overestimation of nonbonding interactions in the DF approach may lead to local minima that are different from those obtained with the IMOMM approach because the DF energetics favor structures involving artificially close contacts. Of course, on a potential energy surface in a high-dimensional space, it is quite difficult with a standard optimization procedure to avoid localizing a metastable local minimum instead of the true ground state. Yet, it is encouraging that, in the example discussed above, the more realistic treatment of van der Waals interactions in the QM/MM approach resulted in the identification of low-lying minima.

Conclusions

We carried out the first QM/MM study on a ligand-protected d metal cluster, using a density-functional-based IMOMM approach. As a model compound, we chose the copper thiolate cluster $Cu_{13}(SCH_2CH_3)_8$, imposing D_{4h} symmetry constraints to enable the evaluation of ligand arrangements with bridge-hollow and top coordination at the Cu_{13} cluster core. We considered various orientations of the ligands relative to the surface of the metal particle as well as staggered and eclipsed forms of the ethyl end group of the ligands, yielding a test set of 16 conformers. Structures have been optimized with an LDA functional; energies were evaluated in single-point fashion using a GGA functional. To assess the performance and accuracy of the IMOMM approach, we optimized all 16 isomers of $Cu_{13}(SCH_2CH_3)_8$ at both the all-electron DF and the QM/MM levels, using one starting geometry for each isomer.

With some exceptions, we found that the QM/MM approach reproduces the results of the full QM calculations in satisfactory fashion, for both structure and energy data. The structures of a pair of staggered and eclipsed rotamers (bdoe/bdos) showed close Cu−H or H−H contacts in the (original) QM structures which were not reproduced in the corresponding QM/MM calculations. In that case, QM and QM/MM results for structure and energetics differed substantially. Less severe differences were found for other isomers exhibiting short nonbonding contacts (buis/buie, tuis/tuie, and tdos/tdoe). In the DF calculations with standard exchange-correlation functionals, such close nonbonding contacts lead to artificial attractive interactions which are absent in QM/MM calculations when van der Waals interactions are handled at the MM level. Reoptimization of all QM structures starting from QM/MM-optimized geometries yielded new QM structures for those three isomers that previously had exhibited the strongest deviations from QM/MM results. The isomers of bdo obtained with the new, full QM calculations exhibit ∼0.8 Å longer Cu−H distances (bdoe 2.76 Å, bdos 3.24 Å). Nevertheless, the H−H contacts increased by only 0.2 Å for bdoe compared to the old isomer. Figure 4 (QM2) reveals that even the ligands approach each

other more closely in the new, full DF structures than in those obtained with the QM/MM treatment. Thus, one obtains overall good agreement between QM and QM/MM calculations. The shortcomings regarding nonbonding interactions displayed by the DF approach, based on a standard exchange-correlation functional, in general, lead to small deviations of structural and energetic results. For certain cases with the largest structural discrepancies, the optimization actually had resulted in two different local minima.

On the basis of this study, we conclude that the IMOMM approach is capable of treating metal cluster compounds comprising extended ligand shells. However, one has to be aware of potential deviations from a full quantum mechanical treatment when short nonbonding contacts occur, especially across the boundary of the QM and MM parts of the model. This issue is particularly crucial if the QM treatment is based on LDA or GGA density functional calculations. Such exchange-correlation functionals are known to fail for nonbonding (dispersion) interactions. Thus, in contrast to other systems, one may expect that a QM/MM approach, employing a suitably parametrized force field, will yield more reliable results for systems involving many van der Waals contacts in the ligand shell or across the boundary of the QM and MM regions.

Acknowledgment. This work was supported by Deutsche Forschungsgemeinschaft, Volkswagen Foundation, and Fond der Chemischen Industrie.

References

(1) Thiel, W. *THEOCHEM* **1997**, *398*−*399*, 1.

(2) Monard, G.; Merz, K. M., Jr. *Acc. Chem. Res.* **1999**, *32*, 904.

(3) Sherwood, P.; de Vries, A. H.; Guest, M. F.; Schreckenbach, G.; Catlow, C. R. A.; French, S. A.; Sokol, A. A.; Bromley, S. T.; Thiel, W.; Turner, A. J.; Billeter, S.; Terstegen, F.; Thiel, S.; Kendrick, J.; Rogers, S. C.; Casci, J.; Watson, M.; King, F.; Karlsen, E.; Sjøvoll, M.; Fahmi, A.; Schäfer, A.; Lennartz, C. *THEOCHEM* **2003**, *632*, 1.

(4) Maseras, F.; Morokuma, K. *J. Comput. Chem.* **1995**, *16*, 1170.

(5) Woo, T. K.; Pioda, G.; Röthlisberger, U.; Togni, A. *Organometallics* **2000**, *19*, 2144.

(6) Woo, T. K.; Margl, P. M.; Deng, L.; Cavallo, L.; Ziegler, T. *Catal. Today* **1999**, *50*, 479.

(7) Lopez, N.; Pacchioni, G.; Maseras, F.; Illas, F. *Chem. Phys. Lett.* **1998**, *294*, 611.

(8) Fischer, D.; Curioni, A.; Andreoni, W. *Langmuir* **2003**, *19*, 3567.

(9) Kerdcharoen, T.; Liedl, K. R.; Rode, B. M. *Chem. Phys.* **1996**, *211*, 313.

(10) Bryce, R. A.; Vincent, M. A.; Hillier, I. H. *J. Phys. Chem. A* **1999**, *103*, 4094.

(11) Kerdcharoen, T.; Morokuma, K. *Chem. Phys. Lett.* **2002**, *355*, 257.

(12) Ryde, U. *J. Comput.-Aided Mol. Des.* **1996**, *10*, 153.

(13) Eichinger, E.; Tavan, P.; Hutter, J.; Parrinello, M. *J. Chem. Phys.* **1999**, *110*, 10452.

(14) Eurenius, K. P.; Chateld, D. C.; Brooks, B. R. *Int. J. Quantum Chem.* **1996**, *60*, 1189.

(15) Kerdcharoen, T.; Birkenheuer, U.; Krüger, S.; Woiterski, A.; Rösch, N. *Theor. Chem. Acc.* **2003**, *109*, 285.

(16) Andres, R. P.; Bielefeld, J. D.; Henderson, J. I.; Janes, D. B.; Kolagunta, V. R.; Kubiak, C. P.; Mahoney, W. J.; Osifchin, R. G. *Science* **1996**, *273*, 1690.

(17) Daniel, M.-C.; Astruc, D. *Chem Rev.* **2004**, *104*, 293.

(18) Chen, S.; Sommers, J. M. *J. Phys. Chem. B* **2001**, *105*, 8816.

(19) Ang, T. P.; Wee, T. S. A.; Chin, W. S. *J. Phys. Chem. B* **2004**, *108*, 11001.

(20) Tempelton, A. C.; Wuelfing, W. P.; Murray, R. W. *Acc. Chem. Res.* **2000**, *33*, 27.

(21) Hostetler, M. J.; Wingate, J. E.; Zhong, C.-Z.; Harris, J. E.; Vachet, R. W.; Clark, M. R.; Londono, J. D.; Green, S. J.; Stockes, J. J.; Wignall, G. D.; Glish, G. L.; Porter, M. D.; Evans, N. D.; Murray, R. W. *Langmuir* **1998**, *14*, 17.

(22) Belling, T.; Grauschopf, T.; Krüger, S.; Mayer, M.; Nörtemann, F.; Staufer, M.; Zenger, C.; Rösch, N. In *High Performance Scientific and Engineering Computing; Lecture Notes in Computational Science and Engineering*; Bungartz, H.-J., Durst, F., Zenger, C., Eds.; Springer: Berlin, 1999; Vol. 8, p 439.

(23) Belling, T.; Grauschopf, T.; Krüger, S.; Nörtemann, F.; Staufer, M.; Mayer, M.; Nasluzov, V. A.; Birkenheuer, U.; Hu, A.; Matveev, A. V.; Shor, A. M.; Fuchs-Rohr, M.; Neyman, K.; Ganyushin, D. I.; Kerdcharoen, T.; Woiterski, A.; Gordienko, A.; Majumder, S.; Rösch, N. *ParaGauss*, version 3.0; Technische Universität München: Munich, Germany, 2004.

(24) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1986**, *7*, 718.

(25) Dapprich, S.; Komaromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. J. *THEOCHEM* **1999**, *461*, 1.

(26) Shor, A.; Rösch, N. Unpublished results.

(27) Nörtemann, F. Dissertation, Technische Universität München, Munich, Germany, 1998.

(28) Matveev, A. V.; Mayer, M. M.; Rösch, N. *Comput. Phys. Commun.* **2004**, *160*, 91.

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

(30) Hayes, D. M.; Barber, M.; Clark, J. H. R. *J. Chem. Soc., Faraday Trans. 2* **1977**, *73*, 1485.

(31) Perram, J. W.; Petersen, H. G.; De Leeuw, S. V. *Mol. Phys.* **1988**, *65*, 875.

(32) Dunlap, B. I.; Rösch, N. *Adv. Quantum Chem.* **1990**, *21*, 317.

(33) Vosko, S. H.; Wilk, L.; Nusair, N. *Can. J. Phys.* **1980**, *58*, 1200.

(34) Ziegler, T. *Chem. Rev.* **1991**, *91*, 651.

(35) Krüger, S.; Seemüller, T.; Wörndle, A.; Rösch, N. *Int. J. Quantum Chem.* **2000**, *80*, 576.

(36) Desmarias, N.; Jamorski, C.; Reuse, F. A.; Khanna, S, N. *Chem. Phys. Lett.* **1998**, *294*, 480.

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

(38) Perdew, J. P.; Wang, Y. *Phys. Rev. B* **1986**, *33*, 8800.

(39) Häberlen, O. D.; Chung, S.-C.; Stener, M.; Rösch, N. *J. Chem. Phys.* **1997**, *106*, 5189.

(40) van Duijnefeldt, F. B. *IBM Res. Rep.* **1971**, RJ945.

(41) Veillard, A. *Theor. Chim. Acta* **1968**, *12*, 405.

(42) Wachters, A. J. H. *J. Chem. Phys.* **1970**, *52*, 1033.

(43) Hay, P. *J. Chem. Phys.* **1977**, *66*, 4377.

(44) Becke, A. D. *J. Chem. Phys.* **1988**, *88*, 2547.

(45) Lebedev, V. I. *Zh. Vychisl. Mat. Mat. Fiz.* **1976**, *16*, 293.

(46) Genest, A.; Krüger, S.; Gordienko, A. B.; Rösch, N. *Z. Naturforsch., B: Chem. Sci.* **2004**, *59b*, 1585.

(47) Holthausen, M. C.; Koch, W. *A Chemist's Guide to Density Functional Theory*; Wiley-VCH: Weinheim, Germany, 2000.

(48) Osinga, V. P.; van Gisbergen, S. J. A.; Snijders, J. G.; Baerends, E. J. *J. Chem. Phys.* **1997**, *106*, 5091.

(49) Wu, Q.; Yang, W. *J. Chem. Phys* **2002**, *116*, 515.

(50) Langreth, D. C.; Dion, M.; Rydberg, H.; Schröder, E.; Hyldgaard, P.; Lundqvist, B. I. *Int. J. Quantum Chem.* **2005**, *101*, 599.

(51) Goddard, W. A., III. Private communication.

(52) Furche, F.; van Voorhis, T. *J. Chem. Phys.* **2005**, *122*, 164106.

(53) Pérez-Jordá, J. M.; Becke, A. D. *Chem. Phys. Lett.* **1995**, *233*, 134.

(54) van Mourik, T.; Gdanitz, R. J. *J. Chem. Phys.* **2002**, *116*, 9620.

(55) Wilson, E. B. *Adv. Chem. Phys.* **1959**, *2*, 367.

CT050202R

JCTC Journal of Chemical Theory and Computation

Time-Dependent Quantum Wave Packet Calculations of Three-Dimensional He − O_2 Inelastic Scattering

Sinan Akpinar, Fahrettin Gogtas,* and Niyazi Bulut

Department of Physics, Faculty Science and Arts, Firat University, Elazig, Turkey

Received February 10, 2005

Abstract: We have studied a three-dimensional time-dependent quantum dynamics of He − O_2 inelastic scattering by using a recently published ab initio potential energy surface. The state-to-state transition probabilities at zero total angular momentum have been calculated in the energy range of 0.12−0.59 eV, and the product rotational distributions are extracted. J-shifting approximation is used to estimate the probabilities for $J > 0$. The integral cross sections and thermal rate constants are then calculated.

1. Introduction

Over the past years, several time dependent quantum wave packet methods were suggested that time dependent quantum approach is quite useful and transparent for studying the dynamics of elementary chemical process, because it allows the direct calculation of observables and shows the possible elementary mechanism.[1] The time-dependent Schördinger equation is initialized from a known quantum state of the system, and the solution of the time dependent Schrödinger equation yields all possible outcomes of interest arising from this initial point. The results for a large range of collision energy can be obtained from a single solution of time-dependent Schrödinger equation.[2] The time dependent approach recently has been used both for two-dimensional and three-dimensional atom−diatom inelastic scattering by many researchers.[3−11]

The He + O_2 may be considered as a prototypical atom−diatom system for low translational energy scattering studies as O_2 is paramagnetic and hence suitable for magnetic trapping method at low energies.[12,13] Therefore, it has been subject to many studies especially concentrated on the rotational alignment and cooling in seeded supersonic expansions of O_2 in He.[14,15] In general, empirical potential energy functions have been employed to investigate this effect. Recently, Groenenboom and Struniewicz[13] have calculated a three-dimensional ab initio ground potential energy surface. Diatomic potential used was constructed from the ab inito calculation and Rydberg-Klein-Rees (RKR) data fitting.[16] Using this full ab inito potential energy surface, the vibrational structure and predissociation dynamics of He + O_2 have been theoretically investigated.[17] Balakrishnan and Dalgarno[18] have carried out the time-independent quantum mechanical calculations to investigate zero temperature quenching rate coefficients for vibrationally and rotationally excited O_2 in collisions with ^{3}He.

In this paper, we discuss a three-dimensional inelastic scattering of He + O_2 by using a grid-based time dependent quantum wave packet method.[2] The paper is organized as follows. In section 2 we discuss the time dependent quantum theory of atom-molecule inelastic scattering. In the last section, the state-to-state inelastic transition probabilities, product rotational distribution, reaction cross sections, and thermal rate constants for the He + O_2 system are discussed.

2. Theory

The method used here is based on the propagation of a state-selected initial wave function in a series of complex Chebychev polynomials and the use of fast Fourier transform,[19] discrete variable representation (DVR),[20] and potentially optimized discrete variable representation techniques[21] for the action of the Hamiltonian operator. The triatomic Hamiltonian operator with total angular momentum $J = 0$ may be written in terms of Jacobi coordinates as

$$\hat{H} = -\frac{\hbar^2}{2\mu}\frac{\partial^2}{\partial R^2} + \frac{\hbar^2 \mathbf{j}^2}{2\mu R^2} + U(R, r, \gamma) + H_{BC}(r) \qquad (1)$$

where R is the distance between the He atom and the center of mass of O_2, r is the O_2 bond length, and γ is the angle between R and r. μ and μ' are corresponding reduced masses, and $\mathbf{j}$ is the rotational angular momentum operator of the O_2 molecule. $H_{BC}(r)$ is the Hamiltonian operator for the diatomic

10.1021/ct050026m CCC: $33.50
Published on Web 11/30/2005

molecule and $U(R, r,\gamma) = V(R, r,\gamma) - V(R = \infty, r, \gamma = 180°)$. As proposed by Kosloff[22] the solution of the time-dependent Schrödinger equation is written in terms of modified complex Chebychev polynomials in the form

$$\psi(R, r, \gamma, t) = e^{-(i/\hbar)(\Delta E/2+V_{min})t}\sum_{n=0}^{N}(2-\delta_{n0}) \times J_n\left(\frac{\Delta E t}{2\hbar}\right)\Phi_n \quad (2)$$

where $\Phi_n = C_n(-i\hat{H}_{norm})\psi(R, r, \gamma, t = 0)$ with $\psi(R, r, \gamma, t = 0)$ being the initial wave function, $C_n(x)$ complex the Chebychev polynomials (CP), $J_n(x)$ the Bessel functions, and ΔE is the magnitude of the entire energy spread of the spectrum of the unnormalized Hamiltonian operator $\hat{H}$. The propagation requires the operation of the $C_n(-i\hat{H}_{norm})$ on ψ. This is performed by using a three-term recursion relation of the Chebychev polynomials

$$\Phi_{n+1} = -2i\hat{H}_{norm}\Phi_n + \Phi_{n-1} \quad (3)$$

The recurrence is started by setting two initial values as $\Phi_o = \psi(R, r, \gamma, t = 0)$ and $\Phi_1 = -i\hat{H}_{norm}\psi(R, r, \gamma, t = 0)$. The initial wave function has three components describing the translational motion of the incoming atom and the vibrational and rotational motions of the target molecule, respectively. The translational wave function has been described in Gaussian form given an initial kinetic energy. The vibrational eigenvalue and eigenfunctions of the diatomic molecule are calculated by solving the time independent Schrödinger equation.[23] The rotational component of the wave function is expressed in associated Legendre polynomials. The action of the Hamiltonian on the wave function in eq 3 is carried out in the following way: Since the potential energy is diagonal in coordinate space, its action on the wave function involves just the multiplication of the values of the potential with those of the wave function at the same spatial grid points. A uniform grid is used for the coordinate R, and the action of the associated kinetic energy operator on the wave packet is evaluated using fast Fourier transforms.[19] The eigenfunctions of the angular kinetic energy operator are known to be the associated Legendre functions $P_j^K(\cos\gamma_l)$ in the general case. For the present application, as J is taken to be zero, normalized Legendre polynomials may be used. Light et al.[20] have discussed the grid or DVR representation based upon a Gauss-Legendre quadrature scheme, and we use this DVR technique for the angular variable γ in the present work. The angular grid points are just the Gauss-Legendre quadrature points. The DVR method allows one to define a transformation matrix which can be used to transform the wave function from the grid (or DVR) representation, in which a value is associated with each grid point $\{\gamma_l, l = 1, 2,.\}$ to a fixed basis representation (FBR) corresponding to an expansion of the wave packet in terms of normalized Legendre polynomials $\hat{P}_j(\cos\gamma_l)$. (Note that $\hat{P}_j$ is used to denote a normalized Legendre polynomial where $\hat{P}_j = \sqrt{(2j+1)/2}P_j$ and P_j is the usual (unnormalized) Legendre polynomial.) The action of the angular part of the kinetic energy operator on the wave function may be easily evaluated when the wave function is expressed as an expansion in Legendre polynomials. This requires the transformation of the wave function from the grid to the FBR representation. The transformation is accomplished by a unitary transformation matrix $\boldsymbol{T}$ defined in terms of normalized Legendre polynomials as $T_{j,l} = \omega_l^{1/2}\hat{P}_j(\cos\gamma_l)$. If an N_γ Gauss-Legendre quadrature scheme is used, then the maximum value of j in the associated fixed basis representation (FBR) is $j_{max} = (N_\gamma - 1)$. In the FBR representation the action of the angular part of the kinetic energy operator on the wave function is accomplished by simply multiplying by $j(j + 1)/2I$. The final DVR wave function is then obtained by carrying out an inverse transformation from the FBR to the grid or DVR representation. This inverse transformation is carried out by using the Hermitian conjugate of the matrix $\boldsymbol{T}$. The action of diatomic Hamiltonian operator is performed by using a potentially optimized discrete variable representation technique which we described in our previous study.[7]

In numerical evaluation for atom−diatom inelastic scattering, the initial wave packet is located in the asymptotic region of entrance channel and propagated on the potential energy surface toward the strong interaction region. In this work we wish to compute state-to-state inelastic scattering probabilities and must therefore follow the development of the wave packet being reflected from the interaction region. Our method of extracting the state-to-state reaction probabilities from the wave packet dynamics requires us to analyze the wave packet as it passes a line in the asymptotic region. To extract the cross section and other observable quantities from the wave packet dynamics, the wave packet is analyzed at each time step by taking cuts through at a fixed value of the scattering coordinate $R = R_\infty$

$$C_{v'j'}(t) = \int_{r=0}^{\infty}(\sum_k \psi(R_\infty, r, \gamma_k, t)P_{j'}(\gamma_k)\omega_k)\phi_{v'j'}(r)dr \quad (4)$$

where ω_k are the weights in Gauss Quadrature formula.[2] The transition probabilities for the production of specific final vibrational−rotational states from a specified initial reactant level are given by[24]

$$P_{jv,v'j'}(E) = \frac{\hbar^2}{\mu\mu'}k_{v'j'}k_{jv}|\frac{A_{v'j'}(E)}{f(k)}|^2 \quad (5)$$

where $A_{v'j'}(E)$ are the Fourier transform of time dependent coefficients ($C_{v',j'}(t)$). k_{jv} is related to total energy E and rovibrational energy states of the diatomic molecule, ϵ_{jv}, by

$$k_{jv} = \left[\frac{2\mu(E-\epsilon_{jv})}{\hbar^2}\right]^{1/2} \quad (6)$$

In applying the time-dependent quantum methods to scattering problems one is always faced with numerical difficulties associated with the reflection of the wave function from the end of the grid. This artificial boundary reflection is due to the discretization of the continuum space by a finite space. Therefore, the wave packet after being analyzed has to be disposed of before reaching the edges of the grid. At present calculations, a negative complex damping potential with aquadratic form has been used at both edges of the grid. For this reason, the normalized Hamiltonian operator is given in eq 3

$$H_{\text{norm}} = H_{\text{norm}} - iA_R\left[\frac{R_d - R}{R_{\max} - R}\right]^2 - iA_r\left[\frac{r_d - r}{r_{\max} - r}\right]^2 \quad (7)$$

where R_d and r_d are the starting points of the complex damping potential, $R_{\max}$ and $r_{\max}$ are the maximum lengths of the grids, and A_R and A_r are the absorbing potential parameters in R and r, respectively. The range of the damping function is limited only in the damping region. That is, when $R < R_d$ the absorbing potential is taken as zero (same reads when $r < r_d$). To prevent any reflection either from the edges of the grid or from the damping potential itself the absorbing potential not to cause any instability, the absorbing potential parameters are optimized as instructed by Vibok and Balint−Kurti.[25] On the other hand, the Bessel functions play a very important role in the convergence of the CP expansion. The number of terms to be used in the CP expansion is set equal to the argument of the Bessel functions which is given in terms of the energy range of the Hamiltonian as $\Delta Et/2\hbar$. Bessel functions decrease exponentially to zero for n values greater than their argument. Therefore, the CP expansion will be unstable if the energy range of the Hamiltonian operator is underestimated. Despite all these precautions we follow the norm of the wave packet at each time step to make sure that the expansion is stable during the propagation. The calculation of the total crosssections requires having the reaction probabilities for all available J values.

$$\sigma_{v,j}(E) = \frac{\pi}{k_{v,j}^2}\sum_{J=0}^{\infty}(2J+1)P_{jv,v'j'}^{J}(E) \quad (8)$$

One approximate way to estimate the reaction probabilities for $J > 0$ is to use the J-shifting method.[26,27] In the J-shifting method the total reaction probabilities for $J > 0$ are calculated by using

$$P^J(E) = P^{J=0}(E - E_{\text{shift}}^J) \quad (9)$$

where $P^{J=0}(E)$ is the accurately computed reaction probability for $J = 0$, at the total energy E, and $P^J(E)$ is the estimated reaction probability for another value of J. The shifting energy is defined as[26,27]

$$E_{\text{shift}}^J = \frac{\hbar^2 J(J+1)}{2\mu R^2} \quad (10)$$

The state-to-state rate constant can be calculated by Boltzmann averaging of the integral crosssection over the collision energy[28]

$$k_{v,j}(E_C) = \frac{d_f}{k_B T}\left(\frac{8}{\pi\mu_{A+BC}k_B T}\right)^{1/2} \times \int dE_C E_C e^{-E_C/k_B T}\sigma_{v,j}(E_C) \quad (11)$$

where d_f is the electronic degeneracy factor,[28] k_B is the Boltzmann constant, and $E_C = E - \epsilon_{v,j}$ is the collision energy.

3. Results and Discussion

In this section, the theory described above was applied to compute rovibrational transition probabilities in inelastic scattering of He + O_2 ($v = 0$, $j = 0$, 1, 2). The potential energy surface has a local minimum of $0.001167a_o$ at a linear geometry, $r = 2.282a_o$, $R = 6.9a_o$. The sattle point with the energy of 0.000368 above the global minimum is located at $R = 6.9a_o$. The coordinate grid used for the propagation covers He − O_2 separations from $2.76a_o$ to $48.3a_o$ and O−O separations from $0.6846a_o$ to $6.846a_o$. 512 evenly spaced grid points were used in the R. The potentially optimized discrete variable representation (DVR) technique is used to set up r grid points and related basis functions, which then allows a compact grid-based matrix representation of diatomic Hamiltonian operator. 32 potentially optimized r grid points were used in the calculations. The maximum value of the rotational quantum number used in the expansion of the wave function (designated $j_{\max}$) was set equal to 60, which allows for several closed channels at the highest energies in the wave packet. The initial wave packet was centered around a He − O_2 separation of $27.57a_o$ and given a kinetic energy of 0.02 eV along the entrance valley. The wave packet had an effective range of kinetic energy from 0.12 to 0.59 eV.

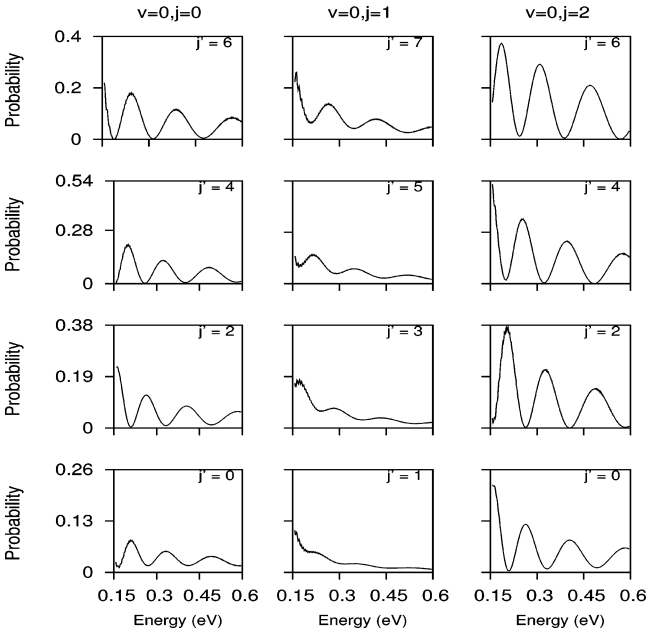


Figure 1. The inelastic transition probabilities for He + O_2($v = 0$, j) → He + O_2($v' = 0$, j') with $j = 0,1,2$.

The time step used for the propagation was approximately 1.2 fs. This small time step leads the wave packet to have a translational energy range of 0.12−0.59 eV. An analysis plane is located at a He − O_2 separation of $34.5a_o$. This plane is defined to lie perpendicularly across the asymptotic region. At each time step, a cut is taken through the wave packet along this plane, and the resulting two-dimensional wave function is analyzed into its fragment state contribution. The analysis of the wave packet as it passes the analysis plane yields the time dependent coefficients. The propagation is continued until all the wave packet has completely left the interaction region, in which case the time dependent coefficients decrease to zero. The portions of the wave packet reflected back into the reaction channel will eventually reach the edge of the numerical grid. If no special precautions are taken, the parts of the wave packet that reach the edge of the grid will be unphysically reflected back onto the grid,

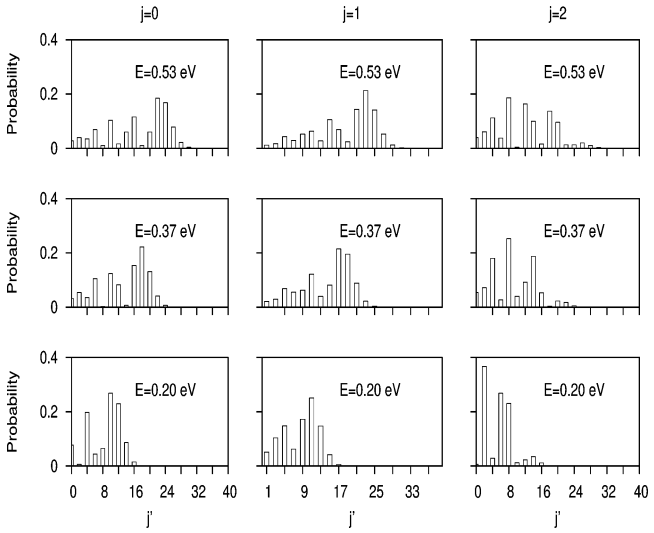


Figure 2. Product rotational distributions at fixed energy values.

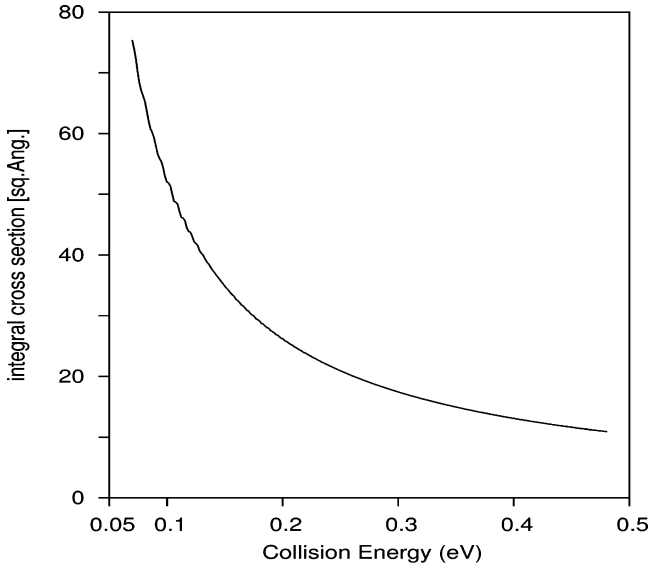


Figure 3. Integral cross sections as a function collision energy for O_2 in its ground.

invalidating the results of the calculations. To avoid such a reflection the damping potential (eq 7) was employed at both edges of the grid at $R_d = 38.58a_o$ and $r_d = 5.22a_o$.

Figure 1 shows the calculated inelastic transition probabilities as a function of translational kinetic energy. As seen from the figure, the individual transition probabilities shows broad oscillatory structures as a function of collision energy. That is, an edge is followed by a monotonic decline. Since the potential energy surface is fully repulsive and has a deep well, this oscillatory feature of the transition probability is attributed to rainbow in the scattering as explained in details by Schinke et al.[29,30] and Levine et al.[31] The state-resolved transition probabilities decrease with increasing energy and have a tendency for even−odd alternation according to the parity selection rule. That is, there is no transition between the even and odd quantum states.

The final rotational distributions for O_2 initially in its ground and first two rotationally excited states are shown in Figure 2 for three different total energy values (0.20, 0.37, 0.53 eV). The rotational distributions show again a rainbow-like structured shape with clear dependency to final rotational quantum number. On the other hand, it may be seen that the shape of the distributions changes as the collision energy increased. That is, the maximum peak of the distribution shifts to higher j' as the translational energy increased, indicating also the energy dependency of rotational distributions.

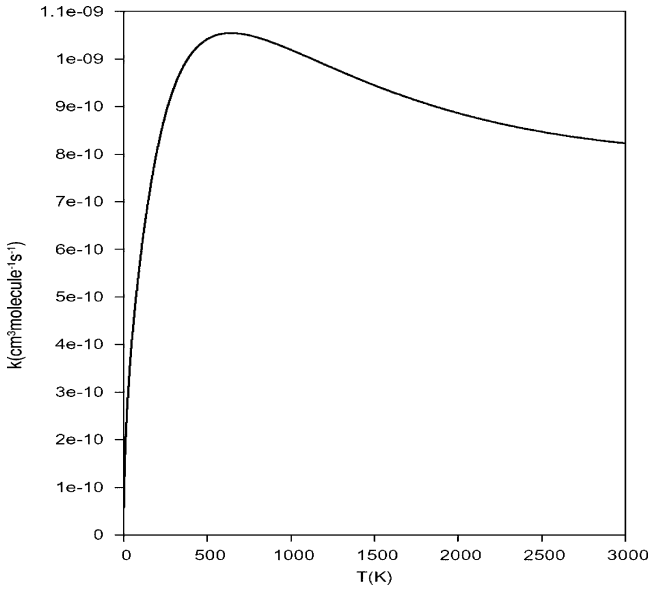


Figure 4. Thermal rate constants derived from the reaction cross sections in Figure 3.

Figure 3 shows the integral cross sections, or excitation functions, corresponding to the O_2 in ground state. The integral cross section has no threshold and is seen to decrease with increasing collision energy. The cross section is very large near zero collision energy and decreases sharply with increase in energy. The thermal rate constants are displayed in Figure 4 for the O_2 in ground state. The rate constants show a weak temperature dependence as it is expected for the reaction with a deep well and high barrier.

References

(1) Defazio, P.; Petrongolo, C. *J. Theor. Comput. Chem.* **2003**, *2*, 547.

(2) Gögtas, F.; Balint-Kurti, G. G.; Offer, A. R. *J. Chem. Phys.* **1996**, *104*, 7927.

(3) Skouteris, D.; Lagana, A.; Capecchi, G.; Werner, H.-J. *Int. J. Quantum Chem.* **2004**, *96*, 547.

(4) Gogtas, F.; Bulut, N.; Oturak, H.; Kökce, A. *J. Mol. Struct. (THEOCHEM)* **2002**, *584*, 149.

(5) Bradley, K. S.; Schatz, G. C.; Balint-Kurti, G. G. *J. Phys. Chem. A* **1999**, *103*, 947.

(6) Akpinar, S.; Gogtas, F.; Bulut, N.; Yildiz, A. *Int. J. Quantum Chem.* **2000**, *79*, 274.

(7) Gogtas, F.; Bulut, N. *Mol. Phys.* **2002**, *100*, 561.

(8) Zhang, D. W.; Wang, M. L.; Zhang, John Z. H. *J. Chem. Phys.* **2003**, *118*, 2716.

(9) Lin, S. Y.; Guo, H. *J. Chem. Phys.* **2003**, *119*, 11602.

(10) Palov, A. P.; Jimeno, P.; Gray, M. D.; Balint-Kurti, G. G. *J. Chem. Phys.* **2003**, *119*, 11602.

(11) Akpinar, S.; Bulut, N.; Gogtas, F. *J. Theor. Comput. Chem.* **2004**, *3*, 291.

(12) Doyle, J. M.; Friedrich, B.; Kim, J.; Patterson, D. *Phys. Rev. A* **1995**, *52*, R2515.

(13) Groenenboom, G. C.; Struniewicz, I. M. *J. Chem. Phys.* **2000**, *121*, 9562.

(14) Aquilanti, V.; Ascenzi, D.; de Castro Vitores, M.; Pirani, F.; Cappeletti, D. *J. Chem. Phys.* **1999**, *111*, 2620.

(15) Fair, J. R.; Nesbitt, D. J. *J. Chem. Phys.* **1999**, *111*, 6821.

(16) Babb, J. F.; Dalgarno, A. *Phys. Rev. A* **1995**, *51*, 3021.

(17) Jung, J.; Sun, H. *Mol. Phys.* **2001**, *99*, 1867.

(18) Balakrishnan, N.; Dalgarno, A. *J. Phys. Chem.* **2001**, *105*, 2348.

(19) Kosloff, R.; Cerjan, C. *J. Chem. Phys.* **1984**, *81*, 3722.

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

(21) Balint-Kurt, G. G.;, Pulay, P. *J. Mol. Struct. (THEOCHEM)* **1995**, *341*, 1−11.

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

(23) Gögtas, F.; Balint-Kurti, G. G.; Marston, C. C. Quantum Chemistry Program Exchange, Program No. 647; QCPE Bulletin, 1994; Vol. 14, p 19.

(24) Balint-Kurti, G. G.; Dixon, R. N.; Marston, C. C. *Faraday Trans. Chem. Soc.* **1990**, *86*, 1741.

(25) Vibok, A.; Balint-Kurti, G. G. *J. Phys. Chem.* **1992**, *96*, 8712.

(26) Bowman, J. M. *J. Phys. Chem.* **1991**, *95*, 4960.

(27) Bittererova, M.; Bowman, J. M. *J. Chem. Phys.* **2000**, *113*, 1.

(28) Lin, S. Y.; Guo, H. *J. Chem. Phys.* **2005**, *122*, 74304.

(29) Schinke, R. *J. Chem. Phys.* **1980**, *72*, 1120.

(30) Mller, W.; Schinke, R. *J. Chem. Phys.* **1981**, *75*, 1219.

(31) Levine, R. D.; Bernstein, R. B. *Molecular Reaction Dynamics and Chemical Reactivity*; Oxford University Press: 1987.

CT050026M

JCTC Journal of Chemical Theory and Computation

Sparkle/AM1 Parameters for the Modeling of Samarium(III) and Promethium(III) Complexes

Ricardo O. Freire,† Nivan B. da Costa Junior,‡ Gerd B. Rocha,† and Alfredo M. Simas*,†

Departamento de Química Fundamental, CCEN, UFPE, 50590-470 - Recife, PE, Brazil, and Departamento de Química, CCET, UFS, 49100−000 - Aracaju, SE, Brazil

Received September 20, 2005

Abstract: The Sparkle/AM1 model is extended to samarium(III) and promethium(III) complexes. A set of 15 structures of high crystallographic quality (*R* factor < 0.05 Å), with ligands chosen to be representative of all samarium complexes in the Cambridge Crystallographic Database 2004, CSD, with nitrogen or oxygen directly bonded to the samarium ion, was used as a training set. In the validation procedure, we used a set of 42 other complexes, also of high crystallographic quality. The results show that this parametrization for the Sm(III) ion is similar in accuracy to the previous parametrizations for Eu(III), Gd(III), and Tb(III). On the other hand, promethium is an artificial radioactive element with no stable isotope. So far, there are no promethium complex crystallographic structures in CSD. To circumvent this, we confirmed our previous result that RHF/STO-3G/ECP, with the MWB effective core potential (ECP), appears to be the most efficient ab initio model chemistry in terms of coordination polyhedron crystallographic geometry predictions from isolated lanthanide complex ion calculations. We thus generated a set of 15 RHF/STO-3G/ECP promethium complex structures with ligands chosen to be representative of complexes available in the CSD for all other trivalent lanthanide cations, with nitrogen or oxygen directly bonded to the lanthanide ion. For the 42 samarium(III) complexes and 15 promethium-(III) complexes considered, the Sparkle/AM1 unsigned mean error, for all interatomic distances between the Ln(III) ion and the ligand atoms of the first sphere of coordination, is 0.07 and 0.06 Å, respectively, a level of accuracy comparable to present day ab initio/ECP geometries, while being hundreds of times faster.

Introduction

Lanthanide complexes and supramolecular architectures have been employed in various areas such as sensors,[1] liquid crystalline materials,[2] electroluminescent devices,[3] luminescent labels for specific biomolecule interactions,[4] and powerful catalysts for various organic transformations.[5] Luminescent lanthanide chelates have been widely used because of their advantages over traditional organic fluorophores: a long decay-time luminescence, large Stokes' shift, narrow emission band, and negligible concentration quenching.

Samarium metal is easily magnetized and difficult to demagnetize.[6] Furthermore, since samarium also has the smallest magnetic moment of all of the paramagnetic lanthanides, it has been used in chiral shift reagents where it presents a greatly reduced line broadening, thereby increasing the reliability of the empirical assignment of the absolute configuration of compounds.[7]

So far, ligand design has mainly produced structures that encapsulate the samarium ion, such as macrocycles and cryptates, creating bulkiness around the metal ion. Since the early 1980s, however, assemblies with two samarium ions facing each other have been discovered[8] and are now appearing in larger numbers.

There is a lack of theoretical methodologies that would permit the a priori design of samarium ligands for various

* Corresponding author tel.: +55 81 2126-8447; fax: +55 81 2126-8442; e-mail: simas@ufpe.br.
† Departamento de Química Fundamental, CCEN.
‡ Departamento de Química, CCET.

10.1021/ct050236z CCC: $33.50
Published on Web 12/08/2005

applications. The ability to efficiently and accurately model all of these samarium molecular systems and interactions is, therefore, an open area of research. More specifically, modeling the influence of the chemical ambience on the $4f^n$ configuration is of significance in the investigation of magnetic and spectroscopic properties of samarium compounds. For example, the description of ligand field effects is central to the design of new ligands capable of forming stable and highly luminescent complexes,[9,10] where the aim is to achieve strong ligand-to-metal energy transfer rates and intense metal-centered emission. The characterization of the interaction between the ligands and the central ion can be done through the ligand field parameters, B_q^k, which can be calculated provided the coordination geometry is known. Within the simple overlap model,[10−12] the values of B_q^k depend mainly on the interatomic distances between the ligand atoms and the central lanthanide ion. This dependence goes with the third, the fifth, and even with the seventh power of the ligand−lanthanide interatomic distances, thus amplifying any inaccuracies. Such interatomic distances are the most sensitive geometric variables impacting upon the description of the effect of the surrounding chemical scenery on the lanthanide ion $4f^n$ configuration. Therefore, a method to accurately predict the geometries of lanthanide complexes from theoretical calculations would be of great advantage. Predicting such geometries may be even more pertinent in light of the fact that obtaining single crystals of lanthanide complexes of appropriate size and optical quality for crystallographic structure determinations may be difficult.[13−16] Reliable, accurate, and fast quantum chemical models for predicting geometries of samarium complexes are urgently needed.

Promethium does not possess any stable isotopes. However, some of them find a variety of uses, such as the activation of zinc sulfide phosphor with β radiation of ^{147}Pm, which provides self-sustaining light sources and is widely used in nocturnal illumination devices.[17] Complexes of promethium radionuclides, mainly ^{147}Pm and ^{149}Pm, have been used in bioresearch, such as in rat age-dependent permeation through skin in vitro,[18] in the development of receptor-based radiopharmaceuticals,[19] and in the radiotherapy of cancer.[20] Radiometals show some significant differences in tumor uptake and retention, physical half-lives, and β-particle path lengths, which may become important determinants of dosimetry and the therapeutic efficacy of pretargeted radioimmunotherapy with these radiolanthanides. The choice of therapeutic radionuclide depends on various factors, such as disease type, stage, and tumor burden; there is not a single ideal radionuclide for cancer therapy. ^{149}Pm, especially, must always be considered as an option because of its α and β energies for the targeted radiotherapy of cancer, low energy, and low-abundance γ emissions, suitable for tracking radiopharmaceuticals in vivo and estimating absorbed radiation doses.[20] Hence, the availability of a fast and accurate a priori quantum chemical model for the prediction of structures of Pm(III) coordination compounds could be of help in the design of promethium complexes exhibiting high thermodynamic, kinetic, and in vivo stabilities.

Ab initio calculations of lanthanide complexes have been sparsely appearing in the literature using various types of effective core potentials, ECPs.[21−27] ECPs replace the chemically inert core electrons of the lanthanide with a potential acting on the valence electrons, which can also be derived to take into account relativistic effects. However, such ECP calculations still demand a large amount of CPU time, rendering high-quality calculations on systems of real chemical interest impractical. Indeed, samarium ab initio/ECP calculations are exceedingly rare.

To make possible the AM1[28] semiempirical calculation of lanthanide complexes, we recognized that the 4f orbitals are contracted toward the nucleus and shielded from fields outside the ion by the outermost 5s and 5p closed shells and introduced the Sparkle model[29] in which we represent the lanthanide ion by a sparkle, that is, by a Coulombic charge of $+3e$ superimposed to a repulsive exponential potential of the form $\exp(-\alpha r)$, which accounts for the size of the ion. We further introduced into the model Gaussian functions in the core−core repulsion energy term to make it compatible with AM1.[30] Recently,[31] we explicitly included sparkle−sparkle core−core interactions to allow the calculation of dilanthanide compounds and defined a new paradigm, Sparkle/AM1, designed to possess geometry prediction accuracies for lanthanide complexes comparable to present day ab initio/ECP calculations, while being hundreds of times faster. Initially, we presented parametrizations for Eu(III), Gd(III), and Tb(III).[31] In the present paper, we extend Sparkle/AM1 to samarium(III) and promethium(III) complexes.

Sparkle/AM1 for Samarium(III)

The parametrization procedure is a nonlinear minimization of an eight-dimension response function. We used a combination of Simplex and Newton−Raphson methods, aimed at finding one of its local minima, which ideally should both be the global minimum and make chemical sense.

The experimental crystallographic structures of the samarium complexes used were all taken from the Cambridge Structural Database 2004.[32−34] The traditional figure of merit for crystal structures is the crystallographic R factor, which provides a measure of how well the refined structure agrees with the experimental model. In the present study, only structures of high quality were considered, that is, structures with R factors less than 5%. For the current work, 15 different structures of complexes for the samarium(III) ion were also considered for parametrization. The response function, F_{resp}, was thus defined as

$$F_{resp} = \sum_{i=1}\left\{\sum_{j=1}[100(R_{i,j}^{CSD} - R_{i,j}^{calcd})]^2 + \sum_{k=1}\left[\frac{2}{3}(\theta_{i,k}^{CSD} - \theta_{i,k}^{calcd})\right]^2\right\} \quad (1)$$

where index i runs over all different complexes, 100 and $^2/_3$ are coefficients taken from the response function originally used to parametrize MNDO,[35] index j runs over all distances (R) of the samarium(III) ion to each of the directly coordinated atoms from the ligands, superscripts CSD and calcd refer to experimental and calculated quantities, and index k runs over all θ angles formed by all combinations of two of the directly coordinated atoms from the ligands

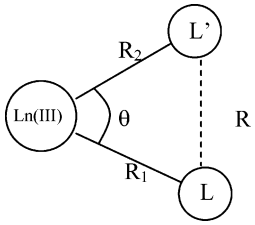


Figure 1. Drawing representing the main variables used in the response function: Ln(III)−L interatomic distances and L−Ln(III)−L′ bond angles. Ln stands for the central lanthanide ion, and L and L′ are ligand atoms of the coordination polyhedron.

Table 1. Number of Samarium(III) and Promethium(III) Complexes in the Validation Sets, Classified into Each Ligand Group by Cluster Analysis

ligand group number	ligand type	number of structures	
		Sm(III)	Pm(III)
1	β-diketones	5	2
2	nitrates	7	2
3	monodentates	7	2
4	bidentates	4	2
5	tridentates	5	2
6	polydentates	8	4
7	dilanthanides	6	1

with the samarium(III) ion in its vertex, as in Figure 1. By adjusting R_1, R_2, and the θ angle, in Figure 1, the L−L′ interatomic distance, which belongs to the coordination polyhedron, was indirectly adjusted.

The next step was to define the set of samarium complex structures to be used in the parametrization procedure, a set we called the samarium parametrization set. It is not simple to classify dozens of structures into smaller representative groups, from which to sample one or two structures to include in the parameter set. Thus, a cluster analysis of all the complexes available in the CSD for samarium was done. The cluster analysis was run with Statistica 6.0 software, using the Euclidean distances with complete linkage to cluster the complexes. As variables, the number of atoms directly coordinated to the lanthanide ion for each of the following types of ligands was used: β-diketone, nitrate, monodentate, bidentate, tridentate, and polydentate. The disamarium complexes were considered a separate group. Only ligands with either or both nitrogen and oxygen as coordinating atoms were considered, since these are the most important ligating atoms for luminescence and most applications.

Moreover, the average unsigned mean error for each complex i, UME_i, was defined as

$$\mathrm{UME}_i = \frac{1}{n}\sum_{j=1}^{n} |R_{i,j}^{\mathrm{CSD}} - R_{i,j}^{\mathrm{calcd}}| \qquad (2)$$

where n is the number of ligand atoms directly coordinating the lanthanide ion.

A set of 15 structures of high crystallographic quality (R factor < 0.05 Å), with ligands chosen to be representative of all complexes in the CSD, with nitrogen or oxygen directly bonded to the samarium ion, was used as a training set (Figure 2). In the validation procedure, we used a set of 42 complexes, also of high crystallographic quality.

Seven molecular groups can be identified from Figure 3. Table 1 describes the molecular clusters and the number of structures found in each.

As previously mentioned, the parametrization procedure used for samarium(III) complexes was identical to the one we successfully used to obtain Sparkle/AM1 parameters for Eu(III), Gd(III), and Tb(III)[31]. The validation procedure has been performed by using, as a measure, the UME, eq 2, this time summing up over all 42 complexes of the validation set.

Table 2. Sparkle/AM1 Parameters for the Sm(III) and Pm(III) Ions

	Sparkle/AM1 Sm(III)	Sparkle/AM1 Pm(III)
GSS	56.993 514 482 0	59.424 970 551 9
ALP	4.175 850 901 0	3.105 983 364 7
a_1	0.959 288 507 0	1.734 767 115 8
b_1	6.479 992 447 0	9.246 422 636 0
c_1	1.738 140 224 0	1.753 341 948 5
a_2	0.026 100 421 0	0.257 101 725 8
b_2	9.739 195 223 0	7.879 344 526 7
c_2	2.888 117 670 0	3.049 816 294 0
EHEAT (kcal mol^{-1})[a]	974.4	976.9
AMS (amu)	150.36	145.0

[a] The heat of formation of the Sm(III) and Pm(III) ions in Sparkle/AM1 was obtained by adding to the heat of atomization of each lanthanide its first three ionization potentials.[38]

Table 3. Values of the Coordination Numbers, CNs, and UMEs for Sparkle/AM1, as Compared to the Respective Experimental Crystallographic Values, Obtained from the Cambridge Structural Database 2004,[32−34] for Each of the 42 Samarium(III) Complexes of the Validation Set

structure[a]	CN	UME (Å) Sparkle/AM1	structure[a]	CN	UME (Å) Sparkle/AM1
ADELAW	7	0.1052	NSMEDT01	9	0.3283
BUVWUK01	9	0.2659	QALFAK	9	0.3233
CAZHAM	8	0.2472	QIHKAT	8	0.1278
CORKEZ	9	0.2588	QIPQOV	9	0.2618
ECABIT	10	0.1187	QOCKIC	8	0.0908
FINDOV	6	0.0541	QQQEMA01	9	0.2129
FUHQOO	9	0.0762	SMNICD	8	0.2563
FUJYEO	8	0.1925	SOXKAR	9	0.2898
GINPEY	9	0.1365	WIGVOX	7	0.1368
GUPHUU	8	0.1523	WOCNIL	4	0.0871
HAWMUN	8	0.1117	XAGVOQ	5	0.0561
JAQNOE	8	0.1312	XAXYAW	7	0.1293
JIZVOD	11	0.1571	XEPLAF	8	0.2742
KIWROX	10	0.1964	XEXJAL	7	0.0683
KUYBAH	9	0.2448	XILGOO	9	0.1034
LIXDUR	9	0.1469	XIVFIR	8	0.1669
LUHFEZ	10	0.0904	XOGYOH	8	0.0997
MEWGOK	9	0.1300	XOWGAR	9	0.2875
MOXJEO	9	0.3295	YENHOO	9	0.2641
NAFKIO	5	0.1905	YUBPAM	8	0.1388
NOWTUO	9	0.1905	ZALDUL	5	0.2466

[a] The structures are identified by their respective codes of reference from the Cambridge Structural Database 2004.[32−34]

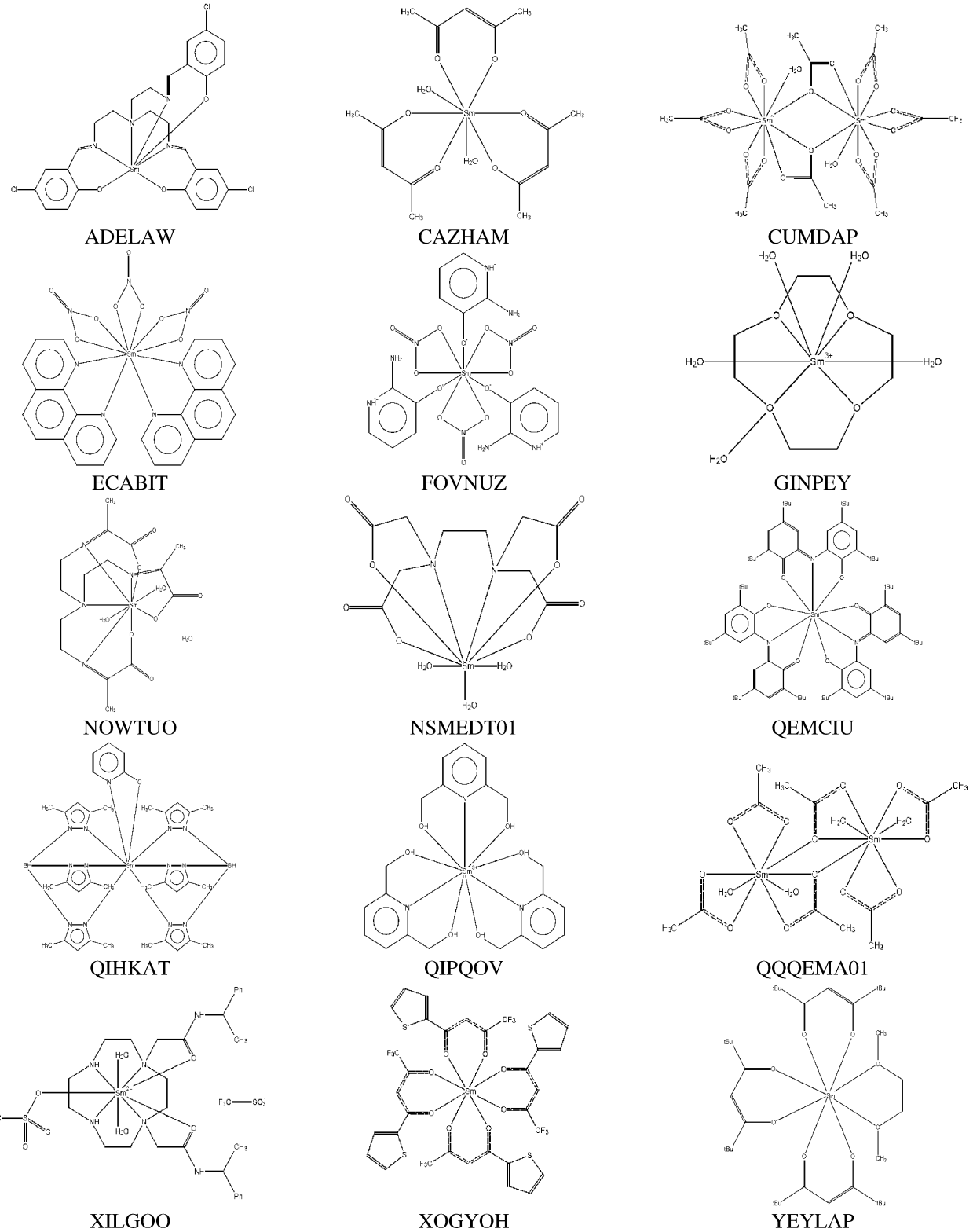


Figure 2. Schematic two-dimensional representations of the samarium(III) complexes that constitute the parametrization training set, obtained from the Cambridge Structural Database 2004.[32−34]

Sparkle/AM1 model calculations have been carried out using the MOPAC93r2 package[36] for the geometry optimization of samarium(III) complexes. MOPAC keywords used in all Sparkle/AM1 calculations were GNORM = 0.25, SCFCRT = 1.D-10 (in order to increase the SCF convergence criterion), and XYZ (the geometry optimizations were performed in Cartesian coordinates).

The best parameter set found that defines the Sparkle/AM1 model for the samarium(III) ion is presented in Table 2.

Our objective, which was to guarantee that Sparkle/AM1 for Sm(III) was as accurate as Sparkle/AM1 for Eu(III), Gd(III), and Tb(III),[31] was achieved. In Table 3, we present the UMEs for all 42 complexes used in the validation test.

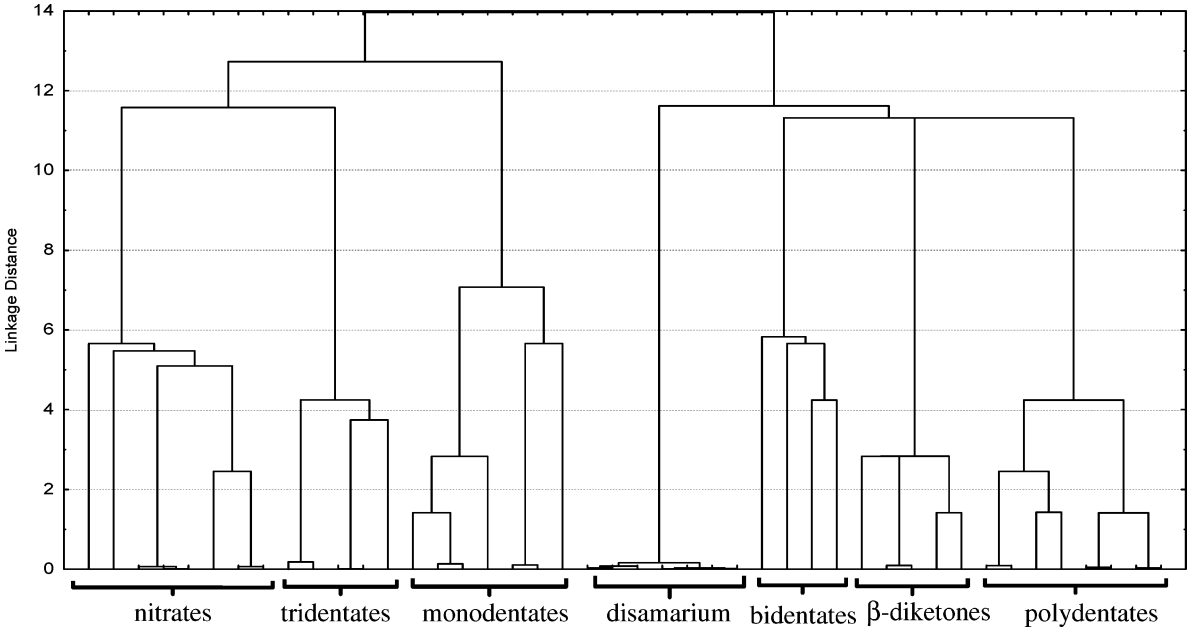


Figure 3. Cluster analysis of all 42 samarium(III) complexes, in terms of both the UMEs and the number of atoms directly coordinated to the lanthanide ion, for each of the various types of ligands. The UMEs are calculated for each complex as the sum of all absolute values of differences between experimental and calculated interatomic distances, involving all atoms of the coordination polyhedron as well as these and the central samarium(III) ion.

If we observe the UMEs in Figure 4 for each of the 42 complexes, grouped according to the cluster analysis shown in Figure 3, where the UME was calculated considering all interatomic distances of the coordination polyhedron, we can see that only three complexes, MOXJEO, NSMEDT01, and QALFAK, present UMEs above 0.3 Å. In the three cases, the high UMEs of 0.330, 0.328, and 0.323 are mainly due to problems in the description of L−L distances, where L is an atom of the ligand directly coordinating the samarium ion. The errors caused by the L−L distances correspond to 82%, 87%, and 92% of the total UMEs, respectively.

However, by analyzing only the distances involving the Sm(III) ion, Figure 4b, we can observe that most of the structures show a UME below 0.20 Å. Table 4 shows UMEs separated by more specific types of bonds and angles. As mentioned before, distances between samarium(III) and oxygen or nitrogen ligand atoms are the most important for the design of luminescent complexes, their UMEs being 0.064 and 0.095, for Sm(III)−O and Sm(III)−N, respectively.

Sparkle/AM1 for Promethium (III)

The fact that there are no crystallographic structures of promethium complexes available, although promethium complexes are being used in bioresearch, makes it even more useful to have a semiempirical model for them, a model that could be of help in their design.

Accordingly, we then decided to investigate the possibility of parametrizing Sparkle/AM1 for promethium from results of ab initio/ECP calculations using only the quasirelativistic ECP for promethium(III) ions, developed by Dolg et al.[22] and implemented in Gaussian 98 as the MWB50 ECP,[37] together with its related [5s4p3d] − GTO valence basis sets. This ECP includes 46 + $4f^n$ electrons in the core, leaving the outermost 11 electrons to be treated explicitly.

Recently,[31] we presented evidence that either enlarging the basis set or including correlation, or both, in the calculations, does not necessarily lead to higher accurate lanthanide complex coordination polyhedron predictions. Actually, in many cases, it even worsened their geometries.[31] Our previous results further indicated that RHF/STO-3G/ECP or RHF/3-21G/ECP results are seemingly equivalent in accuracy,[31] when we compare MWB52 ECP calculations carried out on seven different Eu(III) complexes.

Therefore, we decided to investigate this fact in greater detail in order to be able to arrive at a reasonable ab initio standard, from which Sparkle/AM1 for promethium could be parametrized. We started with one of the simplest samarium complexes, the isolated cation of nona-aqua-samarium(III) tris(trifluoromethanesulfonate), $[Sm(H_2O)_9]^{3+}$, of CSD code BUVWUK01 (Figure 5), and concentrated on determining which ab initio model chemistry with the MWB51 ECP would more accurately predict its coordination polyhedron only. First, we carried out a series of RHF calculations with basis sets of increasing size. Our results are presented in the two top graphs of Figure 6. Both the UME of the whole coordination polyhedron and the $UME_{(Sm-L)}$ of the samarium ion ligand distances only considerably worsened as the basis set increased. Actually, these errors more than doubled by going from STO-3G to 6-31G*.

We then decided to fix the basis set at STO-3G and studied the effect of improving the model by the inclusion of electron correlation, both by means of the B3LYP functional and by many-body perturbation theory at the MP2 level. Again, by adding correlation, the predicted coordination polyhedron became worse, as can be clearly seen in the two graphs in

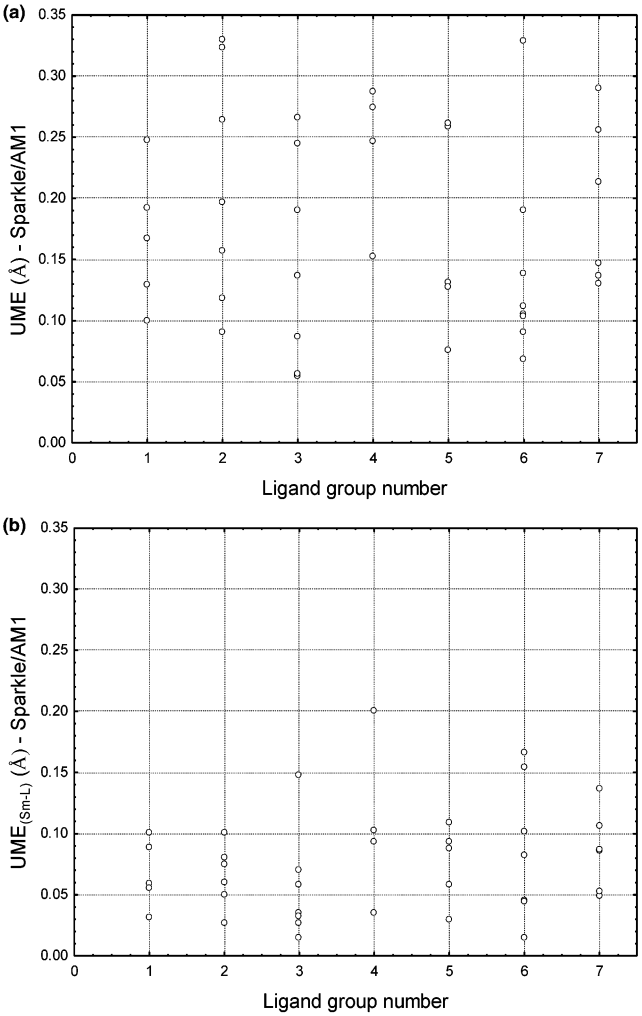


Figure 4. Unsigned mean errors for each of the 42 Sm(III) complexes, assembled according to the ligand group numbers defined in Table 1. Part a presents the UMEs, and part b presents the $UME_{(Sm-L)}$ values. The same scale has been used in both parts to facilitate comparison.

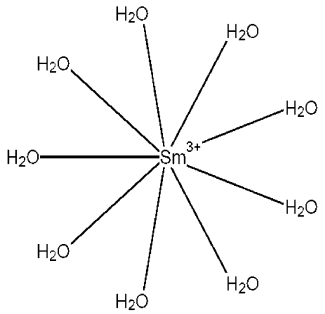
the middle of Figure 6. However, the effect of including electron correlation was much smaller than the basis set effect and affected the whole coordination polyhedron more than the europium−ligand atom distances only. Finally, we decided to examine both factors together, by increasing the basis set size together with including electron correlation. The same trend cropped up, and the two bottom graphs of Figure 6 confirm that RHF/STO-3G is seemingly the most accurate ab initio model for lanthanide complex coordination polyhedron crystallographic geometry prediction from isolated lanthanide complex ion calculations.

We then decided to confirm that RHF/STO-3G/ECP full geometry optimizations of a few representative samarium complexes of known crystallographic geometries would yield coordination polyhedra errors comparable to those of Sparkle/AM1 for the same complexes, thus justifying the use of promethium RHF/STO-3G/ECP calculations to obtain Sparkle/AM1 parameters for promethium. Figure 7 presents the seven samarium complexes chosen, one from each of the clusters of Figure 3, including the largest of all, the disamarium complex of CSD code MEWGOK, with 116 atoms.

Table 4. Values of the Coordination Numbers, CNs, andUMEs for Each of the 15 Promethium(III) Complexes of the Validation Set, for Sparkle/AM1, as Compared to Their Respective Fully Optimized RHF/STO-3G/ECP Geometries[a]

structure[1]	CN	UME (Å) Sparkle model
BUVWUK01{Pm}	9	0.1612
CAZHAM{Pm}	8	0.2043
FINDOV{Pm}	6	0.0826
FUHQOO{Pm}	9	0.1582
FUJYEO{Pm}	8	0.1548
GUPHUU{Pm}	8	0.1276
KUYBAH{Pm}	9	0.2133
LUHFEZ{Pm}	10	0.1422
NOWTUO{Pm}	9	0.1389
NUQYUT{Pm}	6	0.0844
QALFAK{Pm}	9	0.2044
QIPQOV{Pm}	9	0.1328
SOXKAR{Pm}	9	0.2594
XEXJAL{Pm}	7	0.1121
XILGOO{Pm}	9	0.1380

[a] These geometries were obtained by using, as starting points, the geometries of the respective samarium complexes obtained from the Cambridge Structural Database 2004[32−34] and by replacing samarium with promethium. For example, XILGOO{Pm} represents the samarium XILGOO complex, with Pm instead of Sm.



BUVWUK01

Figure 5. Schematic representation of the structure of the cation nona-aqua-samarium(III) obtained from the Cambridge Structural Database 2004.[32−34]

Figure 8 shows UMEs and the $UME_{(Sm-L)}$ for these complexes for both Sparkle/AM1 and RHF/STO-3G/ECP, where one can clearly see that the errors are comparable and trend similarly across the complexes. The results indicate that the Sm(III) parametrization of the Sparkle/AM1 model is capable of predicting coordination polyhedra for most structures with an accuracy equivalent to that of ab initio RHF/STO-3G/ECP. Only for the complex with a polydentate ligand, XEXJAL, was the ab initio RHF/STO-3G/ECP methodology more accurate than the Sparkle/AM1 model. However, the coordination polyhedron UME obtained by Sparkle/AM1 and ab initio RHF/STO-3G/ECP are both low and very close: 0.075 Å and 0.079 Å, respectively.

Consider the ratios in CPU time spent in the complete geometry optimization of the seven structures selected for this analysis between ab initio RHF/STO-3G/ECP and Sparkle/AM1 model calculations. These ratios indicate how fast the Sparkle/AM1 calculation is when compared to the

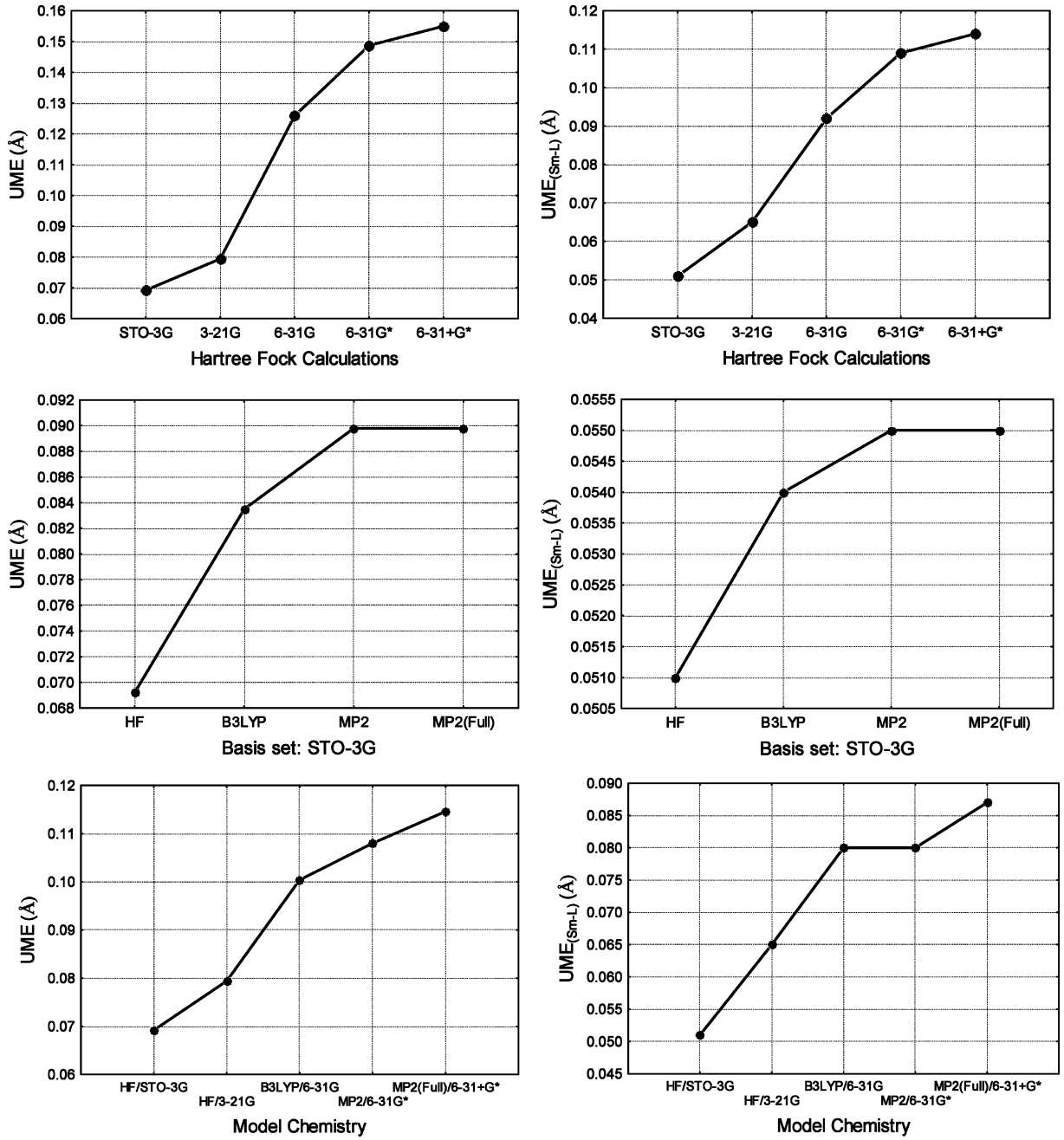


Figure 6. UMEs and $UME_{(Sm-L)}$ for the cation nona-aqua-samarium(III) for various model chemistries, all using the quasirelativistic ECP of Dolg et al.,[22] and all compared to the Cambridge Structural Database 2004[32−34] crystallographic geometry. $UME_{(Sm-L)}$ considers only samarium−ligand atom distances, and UME further includes all distances within the coordination polyhedron.

ab initio one. All have been performed on a Pentium IV 3.0 GHz computer with 2 GB of RAM memory (DDR- 400). For this set of complexes, Sparkle/AM1 calculations ranged from 26 s up to 19 min and were from 115 to 1839 times faster than the corresponding ab initio calculations.

Presently, we do not know the root cause of why RHF/STO-3G/ECP using MWB ECP appears to be the most efficient ab initio model chemistry for coordination polyhedron crystallographic geometry predictions from isolated lanthanide complex ion calculations. But, fortunately, that is so, because the usage of RHF/STO-3G/ECP with MWB ECPs leads to relatively fast ab initio calculations. This finding is warranted only for predictions of coordination polyhedron crystallographic geometries of lanthanide complexes using MWB ECPs. Thus, we cannot assume that this finding would hold true, either for other ECPs or for the geometries of the remaining parts of the molecule. And most likely, this finding will not hold true for the prediction of other properties. However, as we already mentioned, coordination polyhedron geometries are the most sensitive geometric variables impacting upon the description of the effect of the surrounding chemical scenery on the lanthanide ion $4f^n$ configuration.

To obtain the 15-complex promethium parametrization set, we followed the procedure previously used for Eu(III), Gd(III), and Tb(III),[31] and for Sm(III), and chose another set of samarium complexes with ligands representative of the seven types shown in Table 1. Then, as starting points, we

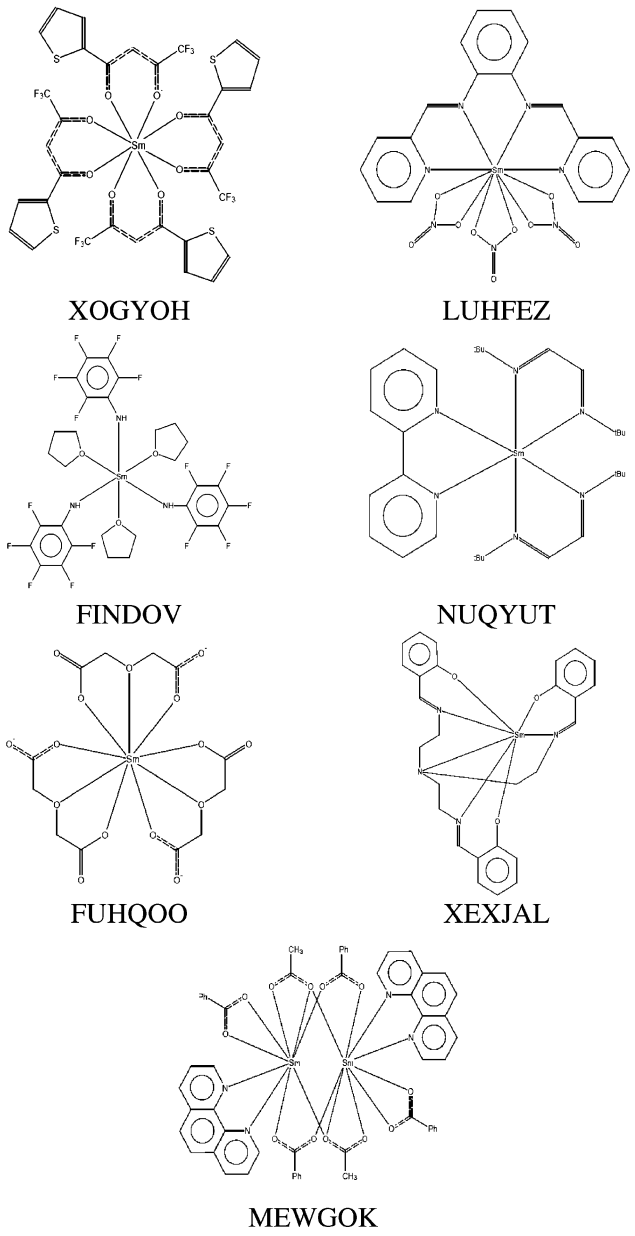


Figure 7. Schematic two-dimensional representations of the structures of samarium(III) complexes, used for comparison between ab initio model geometries and their crystallographic counterparts, identified by their respective Cambridge Structural Database 2004[32−34] codes. The ab initio calculations have been performed using the Hartree−Fock method with the STO-3G basis set for all atoms, except for the samarium(III) ion, in which case we used the quasirelativistic ECP of Dolg et al.[22]

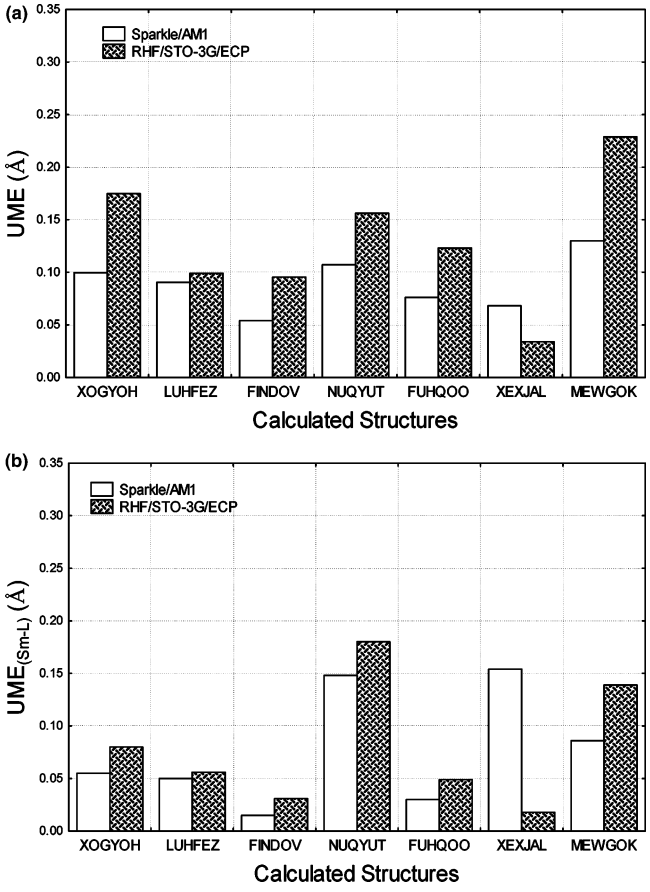


Figure 8. Unsigned mean errors, obtained from Sparkle/AM1 and ab initio RHF/STO-3G/ECP calculations of the ground-state geometries, for each of the seven representative samarium(III) complexes, identified by their respective Cambridge Structural Database 2004[32−34] codes and using the quasirelativistic ECP of Dolg et al.[22] (a) UME includes all distances within the coordination polyhedron and central samarium ion, and (b) $UME_{(Sm-L)}$ considers only samarium−ligand atom distances.

used the geometries of these samarium complexes, replaced samarium with promethium, and fully optimized the geometries with RHF/STO-3G/ECP. We defined a special code for the promethium parametrization set: XILGOO{Pm}, for example, would be the samarium XILGOO complex with Pm instead of Sm. Figure 9 shows the 15 complexes of the promethium parametrization set. The best parameter set found that defines the Sparkle/AM1 model for the promethium(III) ion is also presented in Table 2. Table 4 presents the errors between the promethium Sparkle/AM1 and RHF/STO-3G/ECP results for the 15-complex set, and Table 5 shows the UMEs broken into more specific types of bonds and angles. The figures clearly indicate that results for promethium are, therefore, comparable to results for europium, gadolinium, terbium, and samarium, and the use of the present parametrization for promethium seems warranted until experimental crystallographic structures appear in the literature. As we already indicated, it is precisely when there are no experimental data that theoretical calculations may prove more useful.

Conclusions

Sparkle/AM1 for samarium(III) predicts both lanthanide−ligand distances and distances involving any two atoms in the coordination polyhedron, at the same level of accuracy of the Sparkle/AM1 models as that for Eu(III), Gd(III), and Tb(III) ions. Besides, Sparkle/AM1 accuracy is competitive with, and sometimes better than, ab initio geometries, while being hundreds of times faster.

In conclusion, our results indicate that, for geometry prediction purposes, the Sparkle/AM1 model for Sm(III) is

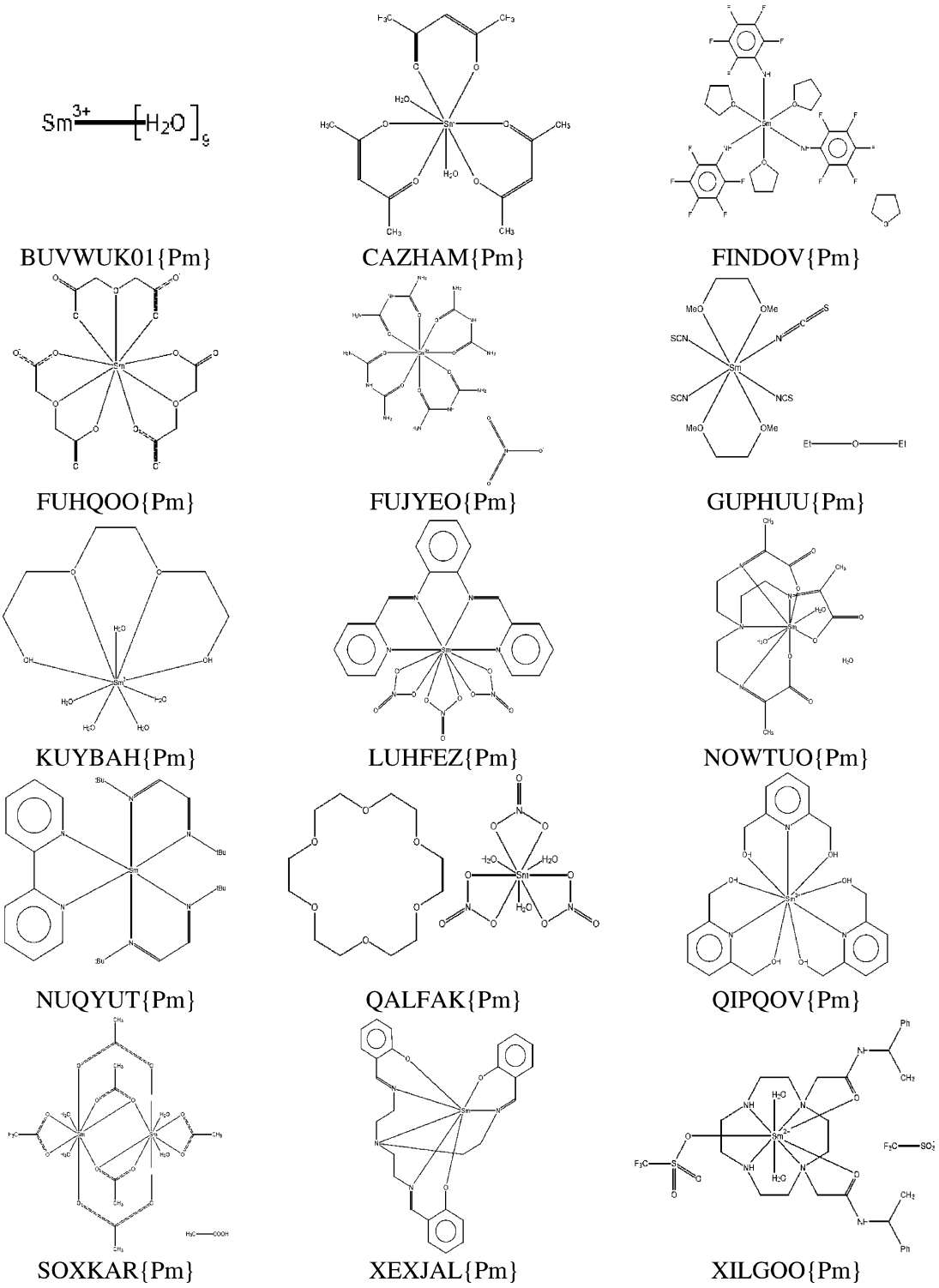


Figure 9. Schematic two-dimensional representations of the promethium(III) complexes that constitute the parametrization training set, obtained from the Cambridge Structural Database 2004.[32−34]

competitive with present day ab initio calculations and may be helpful as a quantum chemical computational technique in the design, for example, of new Sm(III) luminescent compounds where accurate lanthanide−ligand distances are essential.

We also produced further evidence that ab initio/ECP coordination polyhedra geometries of lanthanide complexes seem to be better predicted by RHF/STO-3G/ECP calculations using the MWB ECP. Once again, by increasing the size of the basis set or by including electron correlation, or both, the geometry of the coordination polyhedron of lanthanide complexes is actually worsened. This confirmation allowed us to propose a workable Sparkle/AM1 model for promethium complexes of similar accuracy. Indeed, although there are no promethium complex crystallographic structures available, our Sparkle/AM1 may be of help since, nevertheless, promethium complexes are being currently used in biomedical research.

Table 5. Sparkle/AM1 Unsigned Mean Errors for All Distances Involving the Central Lanthanide Ion, Ln, and the Ligand Atoms of the Coordination Polyhedron, L, for 96 Eu(III) Complexes, 70 Gd(III) Complexes, 42 Tb(III) Complexes, All 42 Sm(III) Complexes, and All 15 Pm(III) Complexes Considered

	unsigned mean errors for specific types of distances (Å)					
model	Ln−Ln	Ln−O	Ln−N	L−L′	Ln−L and Ln−Ln	Ln−L, Ln−Ln, and L−L′
Sparkle/AM1, Eu	0.1624	0.0848	0.0880	0.2170	0.0900	0.1900
Sparkle/AM1, Gd	0.1830	0.0600	0.0735	0.2082	0.0658	0.1781
Sparkle/AM1, Tb	0.2251	0.0754	0.0440	0.2123	0.0746	0.1823
Sparkle/AM1, Sm	0.1381	0.0644	0.0960	0.2158	0.0745	0.1851
Sparkle/AM1, Pm[a]	0.3376	0.0561	0.0591	0.1977	0.0589	0.1681

[a] For promethium, instead of crystallographic geometries as a reference, we used fully optimized RHF/STO-3G/ECP geometries as described in the text.

Acknowledgment. The authors acknowledge the financial support of CNPq, CAPES, FACEPE (Brazilian agencies), PRONEX, Instituto do Milênio de Materiais Complexos, and "Programa Primeiros Projetos" for financial support and CENAPAD (Brazilian institution) for having made available to us their computational facilities. Finally, we gratefully acknowledge the Cambridge Crystallographic Data Centre for the Cambridge Structural Database 2004.

Supporting Information Available: Instructions and examples on how to implement the Sm(III) and Pm(III) Sparkle/AM1 model in Mopac93r2. Parts of the codes of subroutines Block.f, Calpar.f, and Rotate.f that need to be changed, as well as their modified versions for Sm(III) and Pm(III). Examples of Mopac93r2 reference geometry input (.dat) and optimized geometry summary output (.arc) files from Sparkle/AM1 calculations for (i) the Sm(III) complexes XAGVOQ and QQQEMA01 and (ii) the Pm(III) complexes XEXJAL{Pm} and SOXKAR{Pm}. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Gameiro, C. G.; da Silva, E. F., Jr.; Alves, S., Jr.; de Sá, G. F.; Santa-Cruz, P. A. *Mater. Sci. Forum* **1999**, *315*, 249.

(2) Binnemans, K.; Görller-Walrand, C. *Chem. Rev.* **2002**, *102*, 2303.

(3) Kido, J.; Okamoto, Y. *Chem. Rev.* **2002**, *102*, 2357.

(4) Thunus, L.; Lejeune, R. *Coord. Chem. Rev.* **1999**, *184*, 125.

(5) Molander, G. A.; Romero, J. A. C. *Chem. Rev.* **2002**, *102*, 2161.

(6) Scholz, W.; Fidler, J.; Schrefl, T.; Suess, D.; Matthias, T. J. *Appl. Phys.* **2002**, *91*, 8492.

(7) Inamoto, A.; Ogasawara, K.; Omata, K.; Kabuto, K.; Sasaki, Y. *Org. Lett.* **2000**, *2*, 3543.

(8) Mital, S. P.; Singh, R. V.; Tandon, J. P. *Synth. React. Inorg. Met.-Org. Chem.* **1982**, *12*, 269.

(9) de Sá, G. F.; Malta, O. L.; Donega, C. M.; Simas, A. M.; Longo, R. L.; Santa-Cruz, P. A.; da Silva, E. F., Jr. *Coord. Chem. Rev.* **2000**, *196*, 165.

(10) Faustino, W. M.; Rocha, G. B.; Silva, F. R. G.; Malta, O. L.; de Sá, G. F.; Simas, A. M. *THEOCHEM* **2000**, *527*, 245.

(11) Malta, O. L. *Chem. Phys. Lett.* **1982**, *87*, 27.

(12) Malta, O. L. *Chem. Phys. Lett.* **1982**, *88*, 353.

(13) Ronda, C. R. *J. Alloys Compd.* **1995**, *225*, 534.

(14) Bünzli, J.-C. G.; Pighet, C. *Chem. Rev.* **2002**, *102*, 1897.

(15) Thompson, L. C. In *Handbook on the Physics and Chemistry of Rare-Earths*; Gschneider, K. A., Eyring, L., Eds.; North-Holland: Amsterdam, 1979.

(16) Mondry, A.; Bukietyñska, K. *J. Alloys Compd.* **2004**, *374*, 27.

(17) Ravi, S.; Mathew, K. M.; Seshadri, N. K.; Subramanian, T. K. *J. Radioanal. Nucl. Chem.* **2001**, *250*, 565.

(18) Kassai, Z.; Koprda, V.; Bauerová, K.; Harangozó, M.; Bendová, P.; Bujnová, A.; Kassai, A. *J. Radioanal. Nucl. Chem.* **2003**, *258*, 669.

(19) Li, W. P.; Smith, C. J.; Cutler, C. S.; Ketring, A. R.; Jurisson, S. S. *J. Nucl. Med.* **2000**, *41* (5), 246.

(20) Lewis, M. R.; Zhang, J. L.; Jia, F.; Owen, N. K.; Cutler, C. S.; Embree, M. F.; Schultz, J.; Theodore, L. J.; Ketring, A. R.; Jurisson, S. S.; Axworthy, D. B. *Nucl. Med. Biol.* **2004**, *31* (7), 973.

(21) Dolg, M.; Stoll, H.; Preuss, H. *J. Chem. Phys.* **1989**, *90*, 1730.

(22) Dolg, M.; Stoll, H.; Savin, A.; Preuss, H. *Theor. Chim. Acta* **1989**, *75*, 173.

(23) Cundari, T. R.; Stevens, W. *J. Chem. Phys.* **1993**, *98*, 5555.

(24) Ross, R. B. *J. Chem. Phys.* **1994**, *100*, 8145.

(25) Dolg, M. In *Modern Methods and Algorithms of Quantum Chemistry*; Grotendorst, J., Ed.; John von Neumann Institute for Computing: Jülich, Germany, 2000; NIC series, Vol. 1, p 479.

(26) Tsuchiya, T.; Nakajima, T.; Hirao, K.; Seijo, L. *Chem. Phys. Lett.* **2002**, *361*, 334.

(27) Cao, X.; Dolg, M. *THEOCHEM* **2002**, *581*, 139.

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

(29) de Andrade, A. V. M.; da Costa, N. B., Jr.; Simas, A. M.; de Sá, G. F. *Chem. Phys. Lett.* **1994**, *227*, 349.

(30) Rocha, G. B.; Freire, R. O.; da Costa, N. B., Jr.; de Sá, G. F.; Simas, A. M. *Inorg. Chem.* **2004**, *43*, 2346.

(31) Freire, R. O.; Rocha, G. B.; Simas, A. M. *Inorg. Chem.* **2005**, *44*, 3299.

(32) Allen, F. H. *Acta Crystallogr., Sect. B* **2002**, *58*, 380.

(33) Bruno, I. J.; Cole, J. C.; Edgington, P. R.; Kessler, M.; Macrae, C. F.; McCabe, P.; Pearson, J.; Taylor, R. *Acta Crystallogr., Sect. B* **2002**, *58*, 389.

(34) Allen, F. H.; Motherwell, W. D. S. *Acta Crystallogr., Sect. B* **2002**, *58*, 407.

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

(36) Stewart, J. J. P. *MOPAC 93.00 Manual*; Fujitsu Limited: Tokyo, Japan, 1993.

(37) 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.7; Gaussian, Inc.: Pittsburgh, PA, 1998.

(38) Lide, D. R. *Handbook of Chemistry and Physics* [CD-ROM]; CRC Press: New York, 2002.

CT050236Z

JCTC Journal of Chemical Theory and Computation

A QTAIM and Electron Delocalization Computational Study of *tert*-Butylmethylene, Trimethylsilylmethylene, and Trimethylgermylmethylene. A New Method for Unambiguously Characterizing the Bonding between Pairs of Atoms in Reaction Intermediates

Daniel A. Poulsen and Nick H. Werstiuk*

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

Received July 1, 2005

Abstract: While studies on the experimental photolytic and thermolytic extrusion of nitrogen from *tert*-butyldiazomethane and *tert*-butyldiazirine and the decomposition of other precursors have shown a mixture of C−H and C−C insertion products depending on conditions, the analogous trimethylsilyldiazomethane undergoes solely Si−C insertion. Description of the singlet *tert*-butylmethylene intermediates potentially involved in the C−H and C−C insertion reactions and were addressed through computational means by Armstrong et al. (*J. Am. Chem. Soc.* **1995**, *117*, 3685−3689). In addition to re-examining singlet *tert*-butylmethylene at a higher level of theory [CCSD/6-311+G(d,p)], we have studied the silicon and germanium analogues trimethylsilylmethylene and trimethylgermylmethylene. A computational atoms-in-molecules and atomic-basin-delocalization-indices analysis established that the singlet carbenes, while exhibiting varying degrees of delocalization, are not bridged species based on the fact that none possess a pentacoordinate methyl group. In addition, from the results, we are able to make a prediction of solely a Ge−C insertion product for the extrusion of nitrogen from trimethylgermyldiazomethane. Most importantly, we demonstrated that a combination of quantum theory of atoms in molecules (QTAIM) molecular graphs, the evaluation of delocalization indices, and a visualization of the closeness of atomic basins—a QTAIM-DI-VISAB analysis—should be considered as the method of choice for unambiguously characterizing the bonding between pairs of atoms not only of carbenes but of other reaction intermediates such as carbocations, carbanions, and radicals.

Introduction

The photolytic and thermolytic extrusion of nitrogen from *tert*-butyldiazomethane and *tert*-butyldiazirine has been experimentally shown to proceed to the C−H insertion product, 1,1-dimethylcyclopropane, and the C−C insertion product, 2-methyl-2-butene, in ratios dependent on the conditions and method of decomposition.[1] While the involvement of excited states of diazirines and diazo compounds can complicate the situation in the photolytic decompositions, it is generally agreed that singlet carbenes, as the first-formed intermediates in thermal processes, can yield insertion and rearrangement products.[2] The complexities of the formation and reactions of *tert*-butylmethylene have been summarized in a paper by Glick and co-workers.[3] The trimethylsilyldiazomethane analogue of *tert*-butyldiazomethane has been reported to undergo similar decomposition, however, only producing the Si−C insertion product trimethylsilene.[4,5] To date, there have been no known experimental reports on the results of decomposition of the Ge analogue.

Various conformations of the carbene *tert*-butylmethylene [:CHC$(CH_3)_3$] and their reactions have been previously

* Corresponding author phone: (905) 525-9140 ext. 23482; e-mail: werstiuk@mcmaster.ca.

10.1021/ct0501654 CCC: $33.50
Published on Web 12/14/2005

Table 1. Energies of Intermediate Carbenes

carbene	Figure	gradient level of theory	CCSD[a]	MP2[b]	ZPE[c]	relative CCSD[d]
syn-tert-butylmethylene	1	MP2/6-31G(d)		−195.661 330	90.180	
syn-tert-butylmethylene	2	CCSD/6-311+G(d,p)	−195.884 009	−195.817 432	88.428	0.000 00
anti-tert-butylmethylene	3	CCSD/6-311+G(d,p)	−195.883 886	−195.814 133	88.310	0.077 18
anti-trimethylsilylmethylene[e]		MP2/6-311+G(d,p)		−446.840 472	83.061	
anti-trimethylsilylmethylene	4	CCSD/6-311+G(d,p)	−446.918 554	−446.840 273	82.927	
anti-trimethylgermylmethylene[e]		MP2/6-311+G(d,p)		−2233.205 448	82.495	
anti-trimethylgermylmethylene	5	CCSD/6-311+G(d,p)	−2233.282 625	−2233.205 302	82.374	

[a] CCSD energy/hartrees. [b] MP2 energy/hartrees. [c] Zero-point energy at 298 K and 1 atm/kcal mol^{-1}. [d] Relative CCSD energy/kcal mol^{-1}. [e] Figures available in the Supporting Information.

investigated through computational methods.[6] However, conclusions about bonding interactions in these intermediates have, until now, been based on the appearance of molecular geometry and not molecular structure. Like other reaction intermediates such as carbocations, carbanions, and radicals, carbenes are prime candidates for analysis by the quantum theory of atoms in molecules (QTAIM)[7] and delocalization and localization indices (DIs and LIs). QTAIM provides a universal indicator of bonding between atoms[8] in the form of a shared interatomic surface with the number of bond paths terminating at the nucleus defining the coordination at an atom and thereby providing an unambiguous definition of bridging. Delocalization between pairs of atomic basins not exhibiting a bond path may also be investigated through the calculation of DIs.[9,10] In our view, the combination of QTAIM molecular graphs, the evaluation of DIs, and a visualization of the atomic basin proximity at isosurface density values in the range of 0.001−0.005 au—the QTAIM-DI-VISAB analysis—should be considered as the method of choice for unambiguously characterizing the bonding between pairs of atoms in transient intermediates and stable molecules.[11]

This computational study presents data that lead to a refinement of the conclusions regarding bonding reached in the previous treatment of *tert*-butylmethylene carbene and presents computational results on the Si and Ge analogues trimethylsilylmethylene [:CHSi(CH_3)$_3$] and trimethylgermylmethylene [:CHGe(CH_3)$_3$]; we report the results of a QTAIM-DI-VISAB analysis of the bonding in these intermediates.

Computational Methods

Singlet carbene geometries were optimized at MP2/6-31G(d), MP2/6-311+G(d,p), MP2/cc-pVTZ, and CCSD/6-311+G(d,p) levels with Gaussian 03.[12] Frequency calculations were made on the resulting stationary points to confirm them as energy minima. Coupled cluster with single and double excitation (CCSD) minima were confirmed with Moeller−Plesset second-order (MP2) frequency calculations using the same basis set. QTAIM analyses of the wave functions to investigate the topology of the electron densities of the optimized intermediates were carried out with AIM2000,[13] and AIMALL97[14] was used to integrate the atomic basins and obtain the atomic overlap matrices required for DI calculations. The program LI-DICALC[9] was used to obtain the DIs.

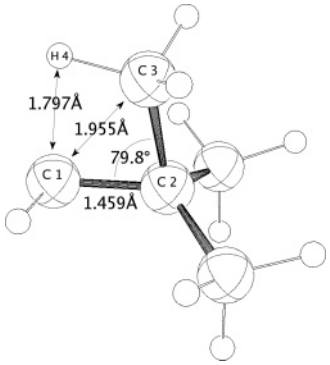


Figure 1. MP2/6-31G(d) *tert*-butylmethylene molecular geometry.

Results and Discussion

We have found that the MP2/6-31G(d) geometry, Figure 1, presented by Armstrong et al.[6] does not exhibit a bond critical point between the carbenic carbon (C1) and the closest methyl carbon (C3) at 1.955 Å, indicating that this minimum-energy geometrical structure should not be considered a bridged species; C3 does not have five bond paths terminating at the nucleus! Nevertheless, the DI between carbenic carbon C1 and C3 (0.278) is higher than the indices between it and both other methyl carbon atoms (0.05), confirming significant electron delocalization afforded by this unsymmetrical geometry. The hydrogen atom H4 also shows a higher DI with C1 than those found on the other two methyl groups; however, no bond critical point exists between them. This intermediate illustrates delocalization without hypervalent bridging, which can be unambiguously defined on the basis of the number of bond paths terminating at the nucleus. Interestingly, we found that this geometry was not an energy minimum using the larger basis sets 6-311+G(d,p) and cc-pVTZ, raising questions as to its validity at the MP2 level.

Calculations at the CCSD/6-311+G(d,p) level yielded two energy-minimum geometries, Figures 2 and 3. These differ in the dihedral angle between the carbenic carbon and the methyl proton. In one, H4 is syn to C1, and in the other, H4 is anti to C1. As seen in the molecular graph, Figures 4 and 5, respectively, no bond critical point or bond path was found between C1 and C3 in either of these geometries and no bond path is seen between H4 and C1 of the syn species. The syn conformation appears similar to the one previously reported at MP2/6-31G(d); however, the C1−C2−C3 angle is larger at 85.2°, implying even less delocalization at this CCSD level. This was confirmed by the atomic basin delocalization analysis, which produced a DI of 0.193

Table 2. Atomic Basin Delocalization Indices

carbene	Figure	level of theory	delocalization index C1−C3	C1−H4	X2−C1[a]	X2−C3[a]
syn-tert-butylmethylene	1	MP2/6-31G(d)	0.278	0.128	0.966	0.737
syn-tert-butylmethylene	2	CCSD/6-311+G(d,p)	0.193	0.103	0.939	0.748
anti-tert-butylmethylene	3	CCSD/6-311+G(d,p)	0.119		0.938	0.763
anti-trimethylsilylmethylene[b]		MP2/6-311+G(d,p)	0.188		0.453	0.399
anti-trimethylsilylmethylene	4	CCSD/6-311+G(d,p)	0.149		0.438	0.400
anti-trimethylgermylmethylene[b]		MP2/6-311+G(d,p)	0.101		0.726	0.646
anti-trimethylgermylmethylene	5	CCSD/6-311+G(d,p)	0.078		0.699	0.643

[a] Where X is C, Si, or Ge. [b] Figures available in the Supporting Information.

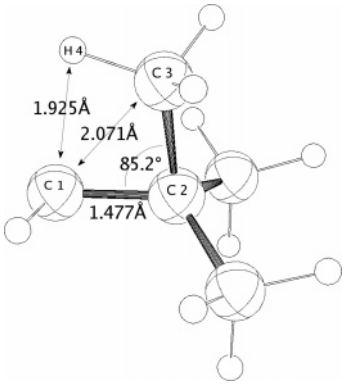


Figure 2. CCSD/6-311+G(d,p) *syn-tert*-butylmethylene molecular geometry.

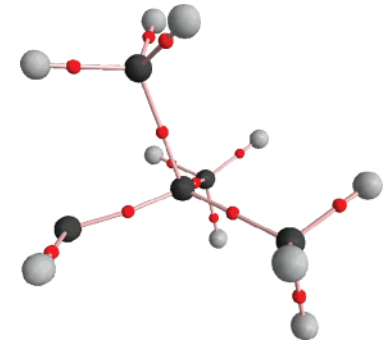


Figure 4. CCSD/6-311+G(d,p) *syn-tert*-butylmethylene molecular graph. The (3,−1) bond critical points are shown as red spheres.

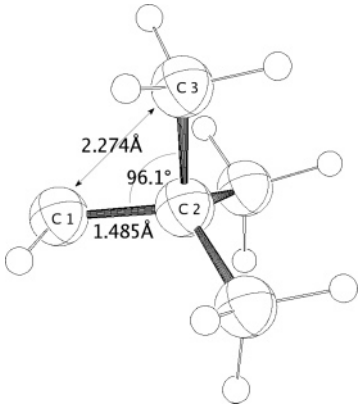


Figure 3. CCSD/6-311+G(d,p) *anti-tert*-butylmethylene molecular geometry.

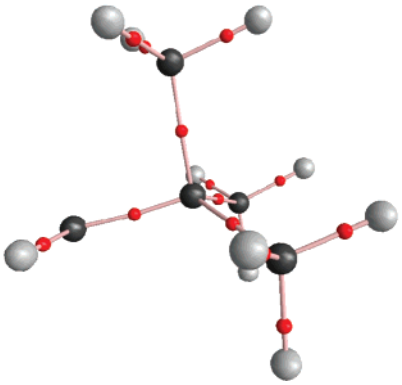


Figure 5. CCSD/6-311+G(d,p) *anti-tert*-butylmethylene molecular graph.

between the C1 and C3 basins compared to 0.278 for the geometry at MP2/6-31G(d). The DI between H4 and C1 of the syn conformer is 0.103 at the CCSD/6-311+G(d,p) level, somewhat less than the value of 0.193 found at MP2/6-31G(d). The DI between H4 and C1 also indicates that significant delocalization, similar to that found between C1 and C3, is not necessarily accompanied by bridging as defined by QTAIM. These carbenes clearly exhibit diffuse electron density which is delocalized into all proximate basins without requiring bridging to each of them.

The difference in energy between the syn and anti species is negligible, with syn more stable than anti by 0.077 kcal/mol. Although both geometries are of essentially equal energy, the delocalization indices between C1 and C3 are quite different at 0.193 and 0.119 for Figures 2 and 3, respectively. This difference is accompanied by minor changes in the delocalization between the central C2 and both the carbenic carbon C1 and methyl group carbon C3. We believe that there are two factors affecting the stability of these intermediates. Delocalization that is stabilizing is the first factor, and it favors both of these unsymmetrical geometries over one with C_s symmetry. The second is destabilizing strain introduced by decreasing the C1−C2−C3 angle. The reason these two geometries are of nearly equal energy is that stabilization achieved in the syn conformation by increased delocalization between C1 and C3 as well as C1 and H4 (DI = 0.103) is countered by the destabilizing effect of decreasing this angle relative to the anti conformation. This analysis indicates that delocalization plays a role in stabilizing these carbenes by favoring the unsymmetrical conformations, but it does not result in the formation of bond paths between the carbenic carbon C1 and C3. These carbenes are far from being bridged species in the QTAIM sense.

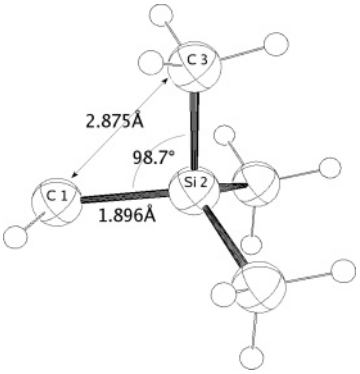


Figure 6. CCSD/6-311+G(d,p) trimethylsilylmethylene molecular geometry.

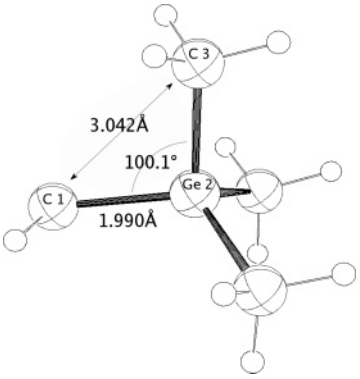


Figure 7. CCSD/6-311+G(d,p) trimethylgermylmethylene molecular geometry.

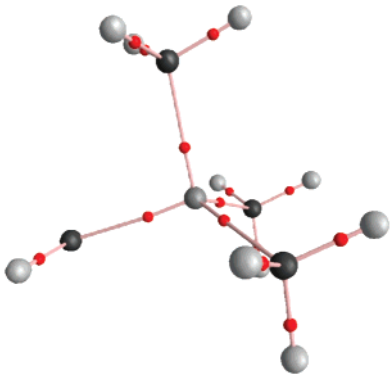


Figure 8. CCSD/6-311+G(d,p) trimethylsilylmethylene molecular graph.

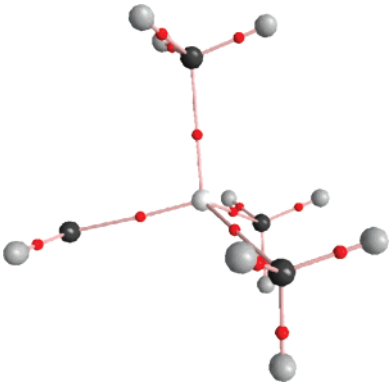


Figure 9. CCSD/6-311+G(d,p) trimethylgermylmethylene molecular graph.

The fact that trimethylsilyldiazomethane does not show C−H insertion upon pyrolysis can be attributed to increased ring strain in the resulting trimethylcylclopropasilane. This is due to a larger Si−C bond distance, making the C−Si−C angle much smaller, at the B3PW91/6-311+G(d,p) level, a C−Si bond length of 1.854 Å and a C1−Si2−C3 angle of 49.8°. The likelihood of the three-member ring is even worse for the germanium analogue, where the carbon−germanium bond distance is even greater, at the B3PW91/6-311+G(d,p) level, a C−Ge bond length of 1.947 Å and C1−Ge2−C3 angle of 46.7°.

We have found similar energy-minimum geometries for the carbenes trimethylsilylmethylene [:CHSi$(CH_3)_3$] and trimethylgermylmethylene [:CHGe$(CH_3)_3$], as seen in Figures 6 and 7, respectively. QTAIM analysis yields no critical point between C1 and C3 in either of these intermediates, as seen in the molecular graphs in Figures 8 and 9. These carbenes exist with H4 of the C3 methyl group anti to the carbenic carbon C1, while the corresponding syn geometry found for *tert*-butylmethylene is no longer an energy minimum for either analogue. The rationale for this, once again, being the effect of increasing the bond length to the central atom. As a result of this distance increase, the delocalization interaction between C1−C3 and C1−H4 is sacrificed owing to the destabilization of the required decrease in the C1−X2−C3 angle. This is consistent with the experimental observations of solely a Si−C insertion product for the decomposition of *tert*-butyldiazomethane.[4,5] The anti geometry is unlikely to undergo C−H insertion with the hydrogen atom out of the plane defined by C1−Si2−C3. Considering the similar geometry of trimethylgermylmethylene, we predict that the extrusion of nitrogen from trimethylgermyldiazomethane will occur in a similar route through Ge−C insertion.

In addition to quantitative results from the QTAIM analysis and atomic basin DIs, we have also plotted several basin surfaces. These plots clearly show the qualitative characteristics of the atomic basin DIs of interest. It is known that an isosurface density value of 0.001 au accounts for 99% of the electronic charge for carbon and can be used to define the van der Waals shape as discussed by Bader.[7] We present plots at an isosurface density value of 0.005 au because the surface is significantly smoother as per limitations of the rendering in AIM2000.[13] In any case, these basins account for even less electronic charge yet still provide a means to visually inspect the relative distances between basins and, therefore, confirm the DI results. Figures 10 and 11 highlight the difference in delocalization for the CCSD/6-311+G(d,p) *syn-tert*-butylmethylene intermediate between the C3 methyl group and the other two. In Figure 10, the C3 basin is in very close proximity with the C1 carbenic basin at an isosurface density value of 0.005 au with a DI of 0.193, while in Figure 11, the C5 basin is significantly separated from C1 with a DI of 0.050. This is consistent with the DI results discussed previously and gives visual confirmation of the basin delocalization for this unsymmetrical intermediate. It is also interesting to note the significant contribution of delocalization between the C1 and H4 basins for the syn orientation. The shape of the H4 basin, shown in Figure 12,

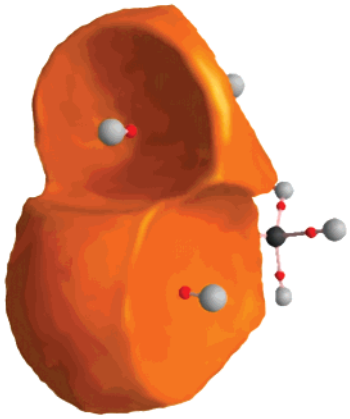


Figure 10. CCSD/6-311+G(d,p) *syn-tert*-butylmethylene molecular graph with atomic basin C1 and C3 density isosurface 0.005.

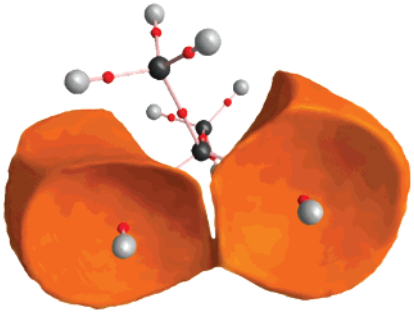


Figure 11. CCSD/6-311+G(d,p) *syn-tert*-butylmethylene molecular graph with atomic basin C1 and C5 density isosurface 0.005.

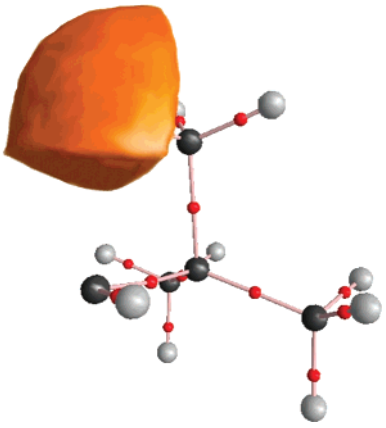


Figure 12. CCSD/6-311+G(d,p) *syn-tert*-butylmethylene molecular graph with atomic basin H4 density isosurface 0.005.

indicates clear deformation caused by the very close proximity of the C1 basin and vice versa, consistent with a DI of 0.103. Figure 10 also clearly illustrates the nearness of the H4 and C1 basins through apparent impingement on each other: Figure 11 clearly shows how the C1 basin is perturbed as a result of the closeness of H4 relative to the C5 basin.

Similar plots of the C1 and C3 basins are presented here for the *anti-tert*-butylmethylene, *anti*-trimethylsilylmethylene, and *anti*-trimethylgermylmethylene intermediates in Figures 13, 14, and 15, respectively. This series highlights the C1 and C3 delocalization difference across the three anti

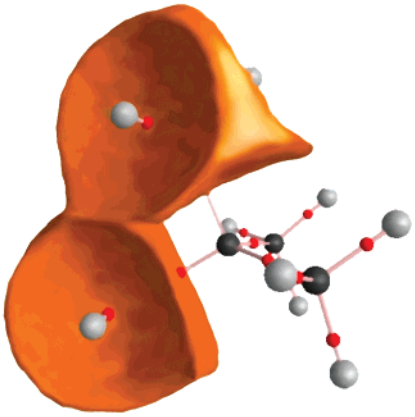


Figure 13. CCSD/6-311+G(d,p) *anti-tert*-butylmethylene molecular graph with atomic basin C1 and C3 density isosurface 0.005.

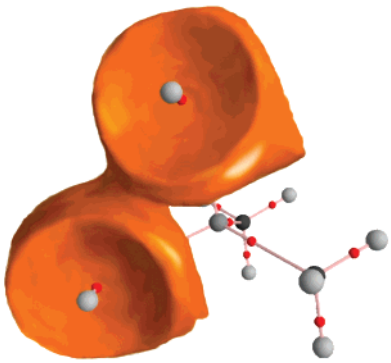


Figure 14. CCSD/6-311+G(d,p) trimethylsilylmethylene molecular graph with atomic basin C1 and C3 density isosurface 0.005.

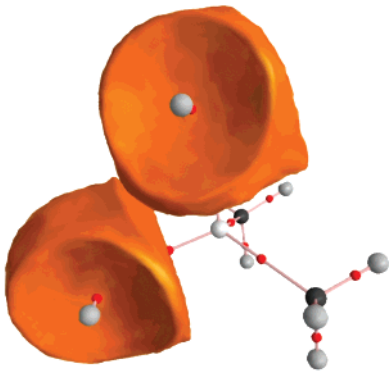


Figure 15. CCSD/6-311+G(d,p) trimethylgermylmethylene molecular graph with atomic basin C1 and C3 density isosurface 0.005.

intermediates which differ in the central atom. At isosurface 0.005, the basins for *anti-tert*-butylmethylene and *anti*-trimethylsilylmethylene are very close to each other, with the DIs being 0.119 and 0.149, respectively, while this is not the case for trimethylgermylmethylene (see Figure 15) with a DI of 0.078.

Conclusions

Using QTAIM, we have shown that free singlet *tert*-butylmethylene does not possess a pentacoordiante methyl

group and, therefore, is not bridged. Two novel geometries have been presented at the CCSD/6-311+G(d,p) level which show a syn and anti conformation with nearly the same energy, $\Delta = 0.077\ 18$ kcal/mol. Delocalization does play a role in stabilizing these carbenes by favoring the unsymmetrical conformations, but they are far from being any sort of bridged species.

Similar intermediate geometries were found for the silicon and germanium analogues, trimethylsilylmethylene and trimethylgermylmethylene; however, the syn conformation was not an energy minimum. This can be attributed to an increase in the Si−C and Ge−C bond lengths. These unsymmetrical carbenes also showed no sign of being bridged species; they do not exhibit pentacoordinate methyl groups. They are stabilized by minor delocalization between a methyl carbon and the carbenic carbon. Our analysis is consistent with experimental findings to date[1,3−5] and is able to predict that the extrusion of nitrogen from trimethylgermyldiazomethane will result in solely the Ge−C insertion product. Most importantly, this paper clearly demonstrates that the combination of QTAIM molecular graphs, the evaluation of DIs, and a visualization of the closeness or atomic basins at isosurface density values in the range of 0.001−0.005 au—a QTAIM-DI-VISAB analysis—should be considered as the method of choice for unambiguously characterizing the bonding between pairs of atoms not only of carbenes but of other reaction intermediates such as carbocations, carbanions, and radicals.

Acknowledgment. We thank the Natural Sciences and Engineering Research Council of Canada for the support making this study possible. We also thank the Shared Hierarchical Academic Research Computing Network (SHARCNET) of Ontario for CPU time necessary for several geometry optimizations.

Supporting Information Available: Figures for *anti*-trimethylsilylmethylene and *anti*-trimethylgermylmethylene at the MP2/6-311+G(d,p) level. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Chang, K. T.; Shechter, H. *J. Am. Chem. Soc.* **1979**, *101*, 5082.

(2) Nickon, A. *Acc. Chem. Res.* **1993**, *26*, 84.

(3) Glick, H. C.; Likhotvorik, I. R.; Jones, M., Jr. *Tetrahedron Lett.* **1995**, *36*, 5715.

(4) Chedekel, M. R.; Skoglund, M.; Kreeger, R. L.; Shechter, H. *J. Am. Chem. Soc.* **1976**, *98*, 7848.

(5) Kreeger, R. L.; Shechter, H. *Tetrahedron Lett.* **1975**, *25*, 2061.

(6) Armstong, B. M.; McKee, M. L.; Shevlin, P. B. *J. Am. Chem. Soc.* **1995**, *117*, 3685.

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

(8) Bader, R. F. W. *J. Phys. Chem. A* **1998**, *102*, 7314.

(9) Wang, Y.; Matta, C.; Werstiuk, N. H. *J. Comput. Chem.* **2003**, *24*, 1720.

(10) Wang, Y.; Werstiuk, N. H. *J. Comput. Chem.* **2003**, *24*, 379.

(11) Bajorek, T.; Werstiuk, N. H. *Can. J. Chem.* **2005**, *83*, 1352.

(12) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; 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*, revisions B.02 and C.02; Gaussian, Inc.: Wallingford, CT, 2004.

(13) Biegler-König, F.; Schönbohm, J. *J. Comput. Chem.* **2002**, *23*, 1489.

(14) Keith, T. A. *AIMAll97 Package (A3) for Windows*; aim@tkgristmill.com.

CT0501654

JCTC Journal of Chemical Theory and Computation

Dependence of the Intermolecular Electrostatic Interaction Energy on the Level of Theory and the Basis Set†

Anatoliy Volkov,* Harry F. King, and Philip Coppens

Department of Chemistry, State University of New York at Buffalo, Buffalo, New York 14260-3000

Received September 2, 2005

Abstract: As electrostatic forces play a prominent role in the process of folding and binding of biological macromolecules, an examination of the method dependence of the electrostatic interaction energy is of great importance. An extensive analysis of the basis set and method dependence of electrostatic interaction energies (E_{es}) in molecular systems using six test dimers of α-glycine is presented. A number of Hartree−Fock, Kohn−Sham, Møller−Plesset, configuration interaction (CI), quadratic CI, and coupled cluster calculations were performed using several double-, triple-, and quadruple-ζ-quality Gaussian- and Slater-type (Kohn−Sham calculations only) basis sets. The main factor affecting E_{es} was found to be the inclusion of diffuse functions in the basis set expansions. Møller−Plesset (even at second order), quadratic CI, and coupled cluster calculations produce the most consistent results. Hartree−Fock and CI methods usually overestimate the E_{es}, while the Kohn−Sham approach tends to underestimate the magnitude of the electrostatic interaction. The combination of the transferable-pseudoatom databank and the exact potential and multipole moment method reproduces Kohn−Sham B3LYP/6-31G** results on which it is based, confirming the excellent transferability of the pseudoatom densities within the systems studied. However, because Kohn−Sham calculations with double-ζ-quality basis sets show considerable deviations from advanced correlated methods, further development of the databank using electron densities from such methods is highly desirable.

Introduction

Electrostatic forces play an important role in the process of protein folding and binding,[1] as the electrostatic interaction energy E_{es} is a major component of the total interaction energy E_{int} of polar molecules. This has long been recognized within the boundaries of the perturbation theory of intermolecular forces[2] in which the electrostatic interaction energy is the leading term in the perturbation expansion of E_{int}:[3]

$$E_{int} = E_{es} + E_{ind} + E_{disp} + E_{ex-rep} \quad (1)$$

where E_{ind}, E_{disp}, and E_{ex-rep} are the induction, dispersion, and exchange-repulsion energies, respectively. E_{es} describes the electrostatic interaction between two unperturbed charge distributions, E_{ind} originates from the interaction of the unperturbed charge density on one monomer with the induced charge distribution on the other (and visa versa), E_{disp} accounts for instantaneous interactions between fluctuating charge distributions on different monomers, and E_{ex-rep} originates from the antisymmetrization of the wave function as a manifestation of the Pauli principle.[4]

We have recently embarked on a quest for an accurate yet efficient evaluation of electrostatic interaction energies in molecular complexes.[5] In widely used force field approaches, E_{es} is commonly calculated with a multipole or Buckingham-type approximation:[2,6]

$$E_{es} = \sum_{i}^{N_A}\sum_{j}^{N_B} \mathbf{T}[\mathbf{r}_{ij}]q_i q_j + \mathbf{T}_{\alpha}[\mathbf{r}_{ij}](q_i\mu_{\alpha,j} - q_j\mu_{\alpha,i}) + \mathbf{T}_{\alpha\beta}[\mathbf{r}_{ij}]\left(\frac{1}{3}q_i\Theta_{\alpha\beta,j} + \frac{1}{3}q_j\Theta_{\alpha\beta,i} - \mu_{\alpha,j}\mu_{\alpha,j}\right) + \ldots \quad (2)$$

* Corresponding author e-mail: volkov@chem.buffalo.edu.

† This paper is dedicated to the memory of Dr. John Rys.

10.1021/ct050216x CCC: $33.50
Published on Web 12/18/2005

where q, μ, Θ, etc. are the permanent atomic moments (monopole, dipole, quadrupole, etc.) in the unperturbed molecular charge distributions and parameters $\mathbf{T}_{\alpha\beta\gamma\ldots}[\mathbf{r}_{ij}]$ are the so-called interaction tensors (with the Einstein summation convention for indices α, β, γ, etc. used), which also depend on the separation of atomic centers $\mathbf{r}_{ij}$. Parameters N_A and N_B represent the number of atoms in molecular fragments A and B, respectively. In many cases, only the first point-charge term of expansion 2 is used,[7−10] although the second and part of the third term of expansion 2 (i.e., charge−dipole and dipole−dipole contributions) have been added in some of the force fields.[11]

In the more advanced distributed multipole approach by Stone and co-workers,[12,13] the expansion is extended to higher-order terms but is still subject to the fundamental limitation of the multipole approximation; that is, it is valid only for nonoverlapping charge distributions. This is especially troublesome for strongly bound systems, involving, for example, short H bonds. In such cases, the multipole approach cannot possibly yield accurate results, and the addition of penetration terms,[12,14] the use of off-atom centered[12] and damping functions,[12] etc. have been proposed. This complicates the calculation process and greatly reduces the transferability of atomic properties.

In our recent paper,[15] we have described a novel approach, called the exact potential and multipole moment (EPMM) method, for the fast and accurate evaluation of electrostatic interaction energies (E_{es}) between two molecular charge distributions within the Hansen−Coppens[16,17] pseudoatom electron density formalism. It combines a numerical evaluation of the exact Coulomb integral for the short-range with the Buckingham-type multipole approximation for the long-range interatomic interactions. It was found, for example, that for intermolecular O···H interactions in molecular systems the multipole approximation underestimates the strength of E_{es}(O···H) by as much as 50 kJ/mol for O···H $\sim$ 1.5 Å, while the EPMM method yields almost that exact result.

We have combined the EPMM method with electron densities from our recently developed theoretical databank of transferable aspherical pseudoatoms,[18,19] referred to below as the DB+EPMM approach. The databank consists of chemically unique pseudoatoms, identified on the basis of common connectivity and bonding. They were extracted from B3LYP/6-31G** densities of a large number of small molecules using a least-squares projection technique in Fourier transform space, and show excellent consistency among chemically equivalent atoms in different molecules. The resulting electrostatic interaction energies E_{es} of monomers in molecular dimers were found to be in a very good agreement with those from a Morokuma−Ziegler decomposition[20,21] of double- and triple-ζ energies[15] evaluated at the density functional level of theory (DFT).

The comparison of E_{es} calculated using the databank parameters (derived from Gaussian-type wave functions) with ADF[22−24] results (in which the Slater-type functions are used and only pure DFT functionals, such as BLYP, are available) is not fully convincing because the two levels of theory used are not equivalent. A meaningful comparison should include intermolecular E_{es} calculated at *exactly* the same level of theory at which the databank parameters were obtained, that is, B3LYP/6-31G**. To this end, a new program, SPDFG, was written for the evaluation of E_{es} from monomer charge distributions expressed in terms of Gaussian-type basis functions. This allows an extensive study of the electrostatic energy of interaction between molecules and its dependence on the orbital basis set for a wide variety of quantum-chemical methods.

Test Systems and Calculations

The current analysis is based on six pairs (dimers) of zwitterionic glycine molecules such as occur in crystals of α-glycine[25] (Figure 1).

Monomer molecular wave functions for Gaussian-type calculations were obtained with the Gaussian03 (G03) suite of programs[26] using methods and basis sets listed in Table 1. The standard Gaussian03 option Output = WFN (and Density = Current for correlated wave functions) generates coefficients of natural orbitals in a primitive basis. For correlated wave functions (MP2, MP4SDQ, CISD, QCISD, and CCSD), generalized densities are based on the Z-vector method.[27−30] All Gaussian03 calculations were performed with the SCF = Tight option, which requests tight self-consistent field convergence criteria.

The new SPDFG program uses the numerical Rys quadrature method[31,32] for the evaluation of one- and two-electron Coulomb integrals. The method is based on a set of orthogonal (Rys) polynomials,[33] which yields a simple general formula for integrals over basis functions, χ, of arbitrarily high angular momentum:

$$\langle\chi_i(1)\chi_j(1)|r_{12}^{-1}|\chi_k(2)\chi_l(2)\rangle = \sum_{\alpha=1}^{N} I_x(u_\alpha)I_y(u_\alpha)I_z^*(u_\alpha)W_\alpha \quad (3)$$

in which u_α and W_α are the roots and weights of the Nth order Rys polynomial and I_x, I_y, and I_z^* are simple two-dimensional integrals, evaluated using efficient and compact recurrence formulas.[32] The program is parallelized using the message-passing interface and can handle basis functions of any angular momentum. E_{es} for Hartree−Fock wave functions evaluated with the SPDFG program are in excellent agreement with those obtained with Morokuma energy decomposition in GAMESS-US.[46]

For Slater-type calculations, E_{es} was obtained using the Morokuma−Ziegler energy decomposition scheme[20,21] implemented in the program ADF,[22−24] which gives electrostatic interaction energies between monomers that are exact within the approximations of the theoretical calculation.

All calculations were performed using our own Linux Beowulf-type cluster equipped with dual-and quad-processor AMD AthlonMP and Opteron nodes.

Results and Discussion

As the electrostatic energy is a major component of the total interaction energy, an analysis of its dependence on the basis set choice and level of theory employed is required for a better understanding of computational results. This is especially important for the evaluation of the performance of the

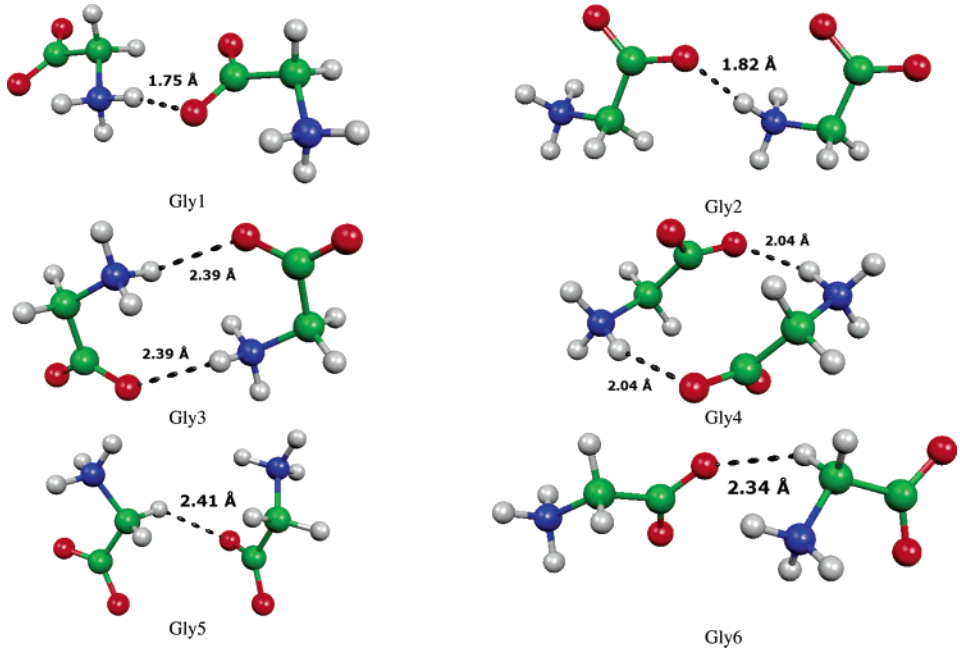


Figure 1. Six dimers in the crystal of α-glycine (oxygen atoms shown in red, nitrogens in blue, carbons in green, and hydrogens in gray).

Table 1. Methods and Basis Sets Used in the Study

methods	basis sets: w/o diffuse functions	basis sets: w/diffuse functions
Gaussian-Type Calculations		
Hartree−Fock (HF)	6-31G**[34]	6-31++G**[35]
DFT with pure BLYP[36] and PBE[37,38] functionals	DZP[39]	DZP+diffuse[40] (DZP+)
DFT with hybrid B3LYP[41] functional	cc-pVDZ[42,43]	aug-cc-pVDZ[42,43]
Møller−Plesset second-order (MP2)	cc-pVTZ[42,43]	aug-cc-pVTZ[42,43]
Møller−Plesset fourth-order with single, double, and quadruple substitutions (MP4SDQ)	cc-pVQZ[42,43,†]	aug-cc-pVQZ[43−45,†]
CI with single and double substitutions (CISD)		
quadratic CI with single and double substitutions (QCISD)		
coupled cluster (CC) with single and double substitutions (CCSD)		
Slater-Type Calculations		
DFT with pure BLYP functional	DZP	
	TZP	
	QZ4P	

† For MP2, HF, and DFT calculations only.

DB+EPMM method, which is to be applied to much larger systems of biological interest to which quantum-mechanical methods are not easily applicable.

1. Effect of Basis Set on the Computed Electrostatic Interaction Energy. *1.1. Comparison of Related Double-, Triple-, and Quadruple-ζ Gaussian and Slater Basis Sets.* The effect of extending the basis set from double-ζ (DZ) to triple-ζ (TZ) is shown in Figure 2a. For Gaussian functions, we report $\Delta E_{es} = E_{es}$(cc-pVTZ) − E_{es}(cc-pVDZ), whereas for Slater functions, TZP and DZP are compared. For Gaussians, the energy calculated with the TZ basis is always more negative (more attractive or slightly less repulsive in the case of dimer Gly5) than the DZ value. The most significant changes are observed for DFT calculations. For example, for Gly3 and Gly4 dimers, ΔE_{es} is as large as 10−15 kJ/mol for pure DFT and 9−11 kJ/mol for hybrid B3LYP functionals. ΔE_{es} at the Hartree−Fock (HF) level is relatively insensitive to the quality of the basis set, the maximum value being just over 4 kJ/mol for the Gly3 dimer. ΔE_{es} values for post-HF calculations are usually intermediate between those for HF and B3LYP.

When considering the dependence of ΔE_{es} on the relative orientation of monomers in dimers, generally, the smallest values are observed for dimers Gly1, Gly2, Gly5, and Gly6 and the largest for dimers Gly3 and Gly4, which are connected by two symmetry-related N−H···O hydrogen bonds.

Slater-type DFT calculations exhibit a different orientational dependence than that observed for Gaussians. For dimer Gly5, the electrostatic energy calculated with the TZ basis is more repulsive by ∼4 kJ/mol than that calculated with the DZ basis. No differences between TZ and DZ basis sets are found for dimer Gly4, in marked contrast to results

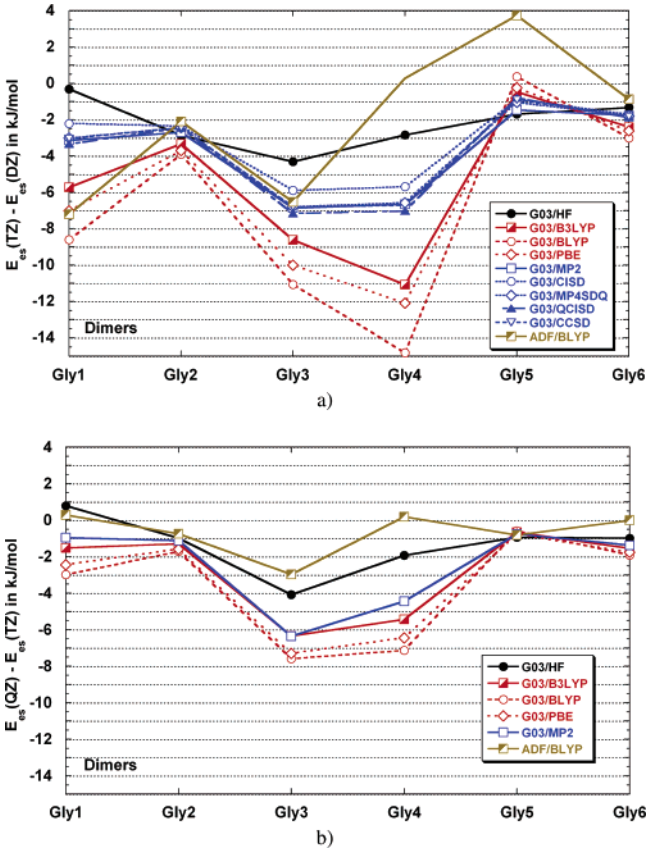


Figure 2. Difference between E_{es} (in kJ/mol) calculated with (a) TZ and DZ basis sets and (b) TZ and QZ basis sets at different levels of theory. For the Gaussian-type calculations, the differences are between (a) cc-pVTZ and cc-pVDZ and (b) cc-pVQZ and cc-pVTZ basis sets. For the Slater-type calculations, (a) TZP-DZP and (b) QZ4P-TZP results are shown.

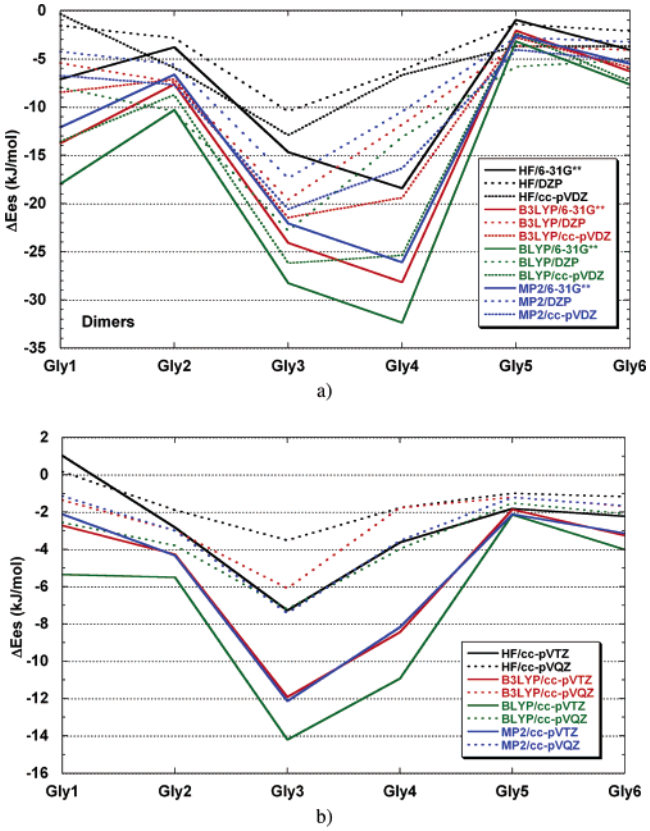


Figure 3. Effect of inclusion of diffuse functions in monomer basis sets on the electrostatic interaction energies in dimers (kJ/mol) for (a) several double-ζ-quality basis sets and (b) triple- and quadruple-ζ basis sets.

obtained with Gaussian functions. The largest deviation (∼7 kJ/mol) is observed for dimers Gly1 and Gly3.

Overall, the differences in E_{es} between the DZ and TZ bases are significant, and in the case of α-glycine dimers can reach 15 kJ/mol. All Gaussian-type calculations show approximately the same dependence of ΔE_{es} on the relative orientation of monomers in dimers, which is different from that observed for DFT calculations with Slater functions.

Figure 2b shows the effect of further expansion of the basis set from triple- to quadruple-ζ (QZ). This leads to corrections for Gaussian DFT and MP2 energies which are smaller than the change between DZ and TZ bases. Although the HF corrections are small, they are comparable to those when going from a DZ to a TZ basis. For Slater-type DFT calculations, the QZ/TZ difference is only −3 kJ/mol for dimer Gly3; −1 kJ/mol for dimers Gly2 and Gly5; and essentially zero for dimers Gly1, Gly4, and Gly6.

For Slater-type calculations, the convergence of E_{es} is nearly complete at the QZ level, while even more extended basis sets are needed to achieve a similar convergence in Gaussian-type calculations; that is, quintuple- or perhaps even sextuple-quality basis sets would be required.

1.2. Effect of Inclusion of Diffuse Functions in the Basis Sets. A prominent result obtained in this study is that the inclusion of diffuse functions in *monomer* charge density calculations has a much more pronounced effect on electrostatic interaction energies than even the change from a simple 6-31G** basis to the cc-pVTZ basis set. Figure 3a shows these effects for DZ-quality basis sets, and Figure 3b shows analogous results for TZ and QZ basis sets. Results are shown only for HF, B3LYP, BLYP, and MP2 calculations, with other methods showing similar behavior.

The inclusion of diffuse functions usually lowers E_{es} (except for a very small positive energy change in HF/cc-pVTZ and cc-pVQZ calculations). Not surprisingly, the 6-31G** basis tends to show a much larger variation in E_{es} upon the inclusion of diffuse functions than any other basis set examined in this study. The change is as small as 2−4 kJ/mol for dimer Gly5 and as large as 28−32 kJ/mol for dimers Gly3 and Gly4. The other two DZ-type basis sets (cc-pVDZ and DZP) are somewhat less affected by the inclusion of diffuse functions than 6-31G**. The maximum changes are ∼25−26 kJ/mol for the cc-pVDZ basis in dimers Gly4 and Gly3 and ∼22 kJ/mol for the DZP basis in dimer Gly3. For dimers Gly2, Gly5, and Gly6, the inclusion of diffuse functions does not significantly affect the E_{es} for any of the DZ-quality basis sets: changes are generally under 10 kJ/mol.

As expected, the effect of including diffuse functions diminishes in going from double- to triple- to quadruple-ζ basis sets. The biggest effects are 7, 14, and 32 kJ/mol for QZ-, TZ-, and DZ-quality basis sets, respectively.

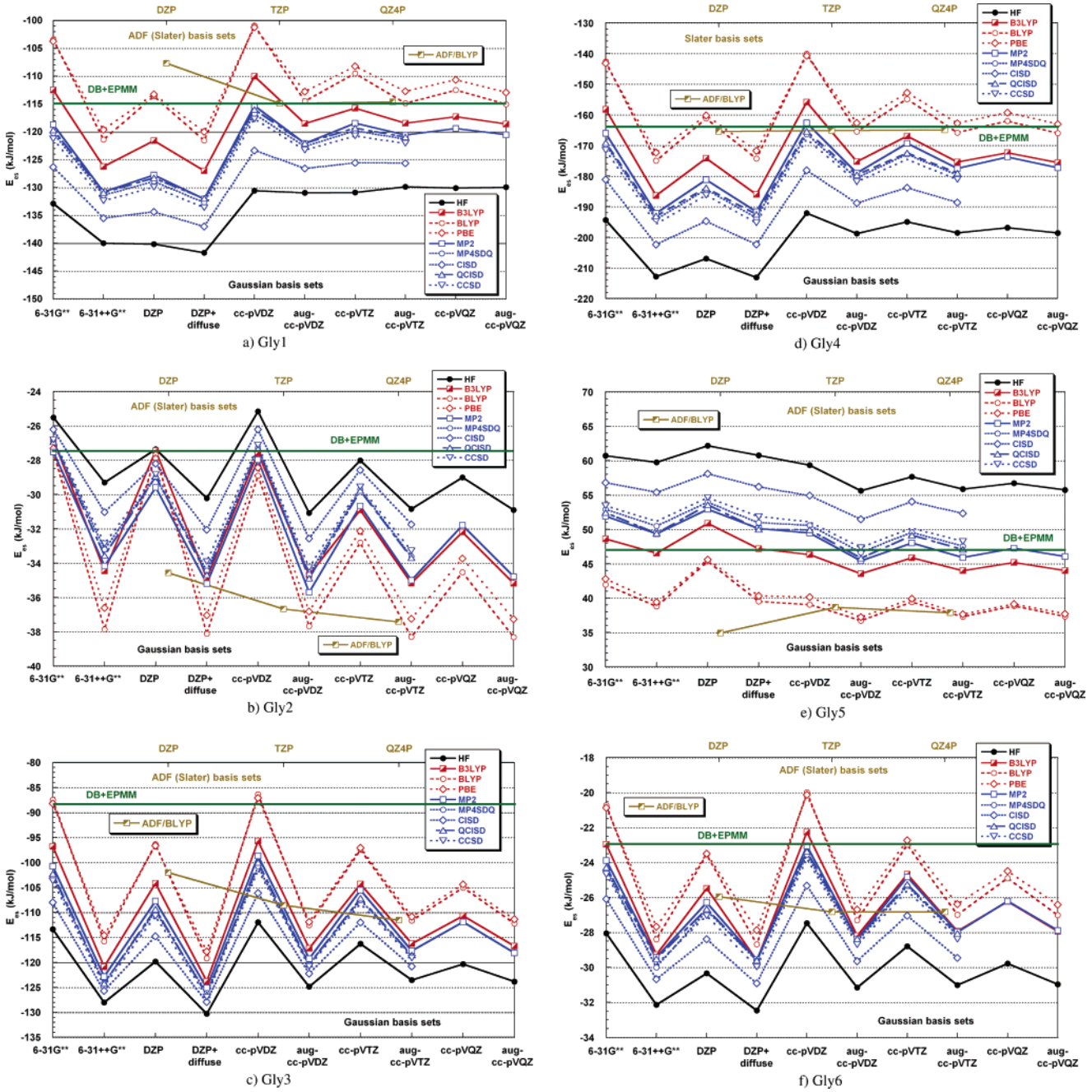


Figure 4. Electrostatic interaction energies (in kJ/mol) in Gly1 (a), Gly2 (b), Gly3 (c), Gly4 (d), Gly5 (e), and Gly6 (f) dimers calculated at different levels of theory.

The inclusion of diffuse functions has its greatest effect on Gaussian DFT (more pronounced for pure DFT functionals) and MP2 energies; is slightly less for CCSD, QCISD, and MP4SDQ; and is the least for HF and CISD methods.

In general, the importance of diffuse functions for the calculation of intermolecular E_{es} reported here is in accord with the results of previous studies, for example, those using symmetry-adapted perturbation theory[3] for various types of systems.[47−49] Similar conclusions were also drawn from the studies of supermolecular interaction energies in both hydrogen-bonded[50,51] and $\pi-\pi$ interacting[52] systems, molecular electric moments and polarizabilities,[53] and so-called correlated cumulative atomic multipole moments.[54] Our results confirm that augmenting a given basis set is more important for the calculation of E_{es}, and also E_{int},[52] than adding a shell of valence functions (i.e., aug-cc-pVDZ vs cc-pVTZ).

It is noteworthy that, within the BLYP method, the electrostatic interaction energies obtained with augmented Gaussian triple- and quadruple-ζ basis sets are in an excellent agreement with those from TZP and QZ4P Slater calculations.

2. Method Dependence of E_{es}. The results summarized in Figure 4a−f show both the basis set and method dependences of E_{es} for each of the α-glycine dimers. The Gaussian-type basis sets are listed at the bottom along the *x* axis, while Slater-type basis sets are listed at the top of the graphs. Values of E_{es} obtained from the DB+EPMM approach are represented by the solid horizontal line.

The overall spread of E_{es} obtained from first principle calculations is remarkably large. For example, it is as large as 74 kJ/mol between pure DFT calculations with cc-pVDZ or 6-31G** basis sets and HF/DZP+ calculations in the Gly4 dimer. Judging from the results of our best calculations, these DFT values are estimated to be ~22% too large and the HF values are ~19% too low. Similarly, the spread of calculated E_{es} values is about 40 kJ/mol in dimers Gly1 and Gly3.

Given the large basis-set dependence, we will analyze the effect of the method on E_{es} within each basis set to arrive at general conclusions. In general, we can distinguish five groups in terms of their E_{es} method dependence: (1) calculations with pure DFT functionals using both Slater and Gaussian functions (BLYP in ADF and BLYP/PBE in G03); (2) hybrid DFT with Gaussian functions (B3LYP in G03); (3) HF; (4) CISD; and (5) MP2, CCSD, QCISD, and MP4SDQ, the last four groups all with Gaussian functions.

For the strongly bonded dimers Gly1, Gly3, and Gly4, all DFT calculations yield less-negative E_{es} values than do the advanced correlated methods, such as CCSD, QCISD, and MP4SDQ. HF, on the other hand, overestimates E_{es} (by 10−20 kJ/mol), as does CISD, but by a smaller amount than HF (~5−10 kJ/mol). The advanced correlated methods, and also MP2, show consistent results with differences of only a few kJ/mol. This is also true for the somewhat more weakly bonded dimer Gly6, except that with aug-cc-pVTZ, and the higher basis set B3LYP method shows an excellent agreement with MP2 results.

For the only repulsive dimer, Gly5, included in this study, the same trend with respect to the method is observed but with the opposite sign, that is, repulsion is largest for HF and more advanced methods.

For the weakly bonded Gly2 dimer, the situation is the opposite of the one described above. HF and CISD calculations underestimate E_{es}, while pure DFT overestimates it. The behavior of the hybrid DFT B3LYP functional in this dimer is similar to that of the MP2, CCSD, QCISD, and MP4SDQ calculations.

Within a given Gaussian basis set approximation, the values of E_{es} either increase or decrease, depending on the spatial orientation of the monomers, in the following order: HF, CISD, (CCSD, QCISD, MP4SDQ), MP2, B3LYP, and pure DFT functionals. As expected, advanced correlated methods, such as CCSD, QCISD, and MP4SDQ, are consistently in good agreement with one another. Electron correlation effects are significant. Large-basis HF calculations yield values that differ from comparable correlated results by factors ranging from 0.95 (Gly1) to 1.19 (Gly5) in reasonable agreement with factors of about 0.94 previously reported for H_2O and HF dimers.[47] MP2 consistently overestimates the magnitude of the electron correlation correction, but never by more than 3 kJ/mol in the six dimers studied here, and these small deviations are removed at the MP4SDQ level of theory. This agrees with early studies of the convergence of the Møller−Plesset perturbation expansion applied to the calculation of electrostatic interaction energies[55] and electron density distributions[56] in simple closed-shell molecules. The CISD method, which suffers from nonsize consistency, recovers only half of the electron correlation correction; that is, CISD E_{es} values are roughly halfway between those of HF and advanced correlated methods. Clearly, CISD is inappropriate for molecules comparable to or larger than glycine.

Hybrid DFT B3LYP calculations often deviate significantly from advanced correlated methods for double-ζ quality basis sets, but the agreement improves for more extended basis sets. The deviations of pure DFT calculations (using either Gaussian or Slater functions) from advanced correlated methods always have the same sign but larger magnitude than those of hybrid B3LYP calculations. Problems with pure DFT functionals have been attributed to their inability to correctly describe long-range correlations,[57,58] which in general can be remedied by incorporation of the special asymptotic correction.[59,60] Hybrid DFT functionals, such as B3LYP, by their very nature, already include a part of correct asymptotics via Hartree−Fock exchange, which improves the overall asymptotic behavior of these functionals. Accordingly, electrostatic energies calculated with pure DFT functionals almost always deviate much more from advanced correlated methods than does hybrid B3LYP. The latter energies are almost always intermediate between those from pure DFT and HF calculations and, in some cases, are even in very good agreement with MP2 results. It is anticipated that, once the asymptotic correction is applied to pure DFT functionals, their performance should improve dramatically and produce electrostatic interaction energies close to those of CCSD.[61]

Several previous studies relate to the method dependence of intermolecular electrostatic interaction energies, either based on the perturbation approach, which adds correlation corrections to the Hartree−Fock E_{es} from perturbation contributions,[47,48,55,62,63] or calculated from relaxed correlated densities.[63] In general, our results, obtained on systems much larger than those studied previously, confirm (a) the importance of intramolecular correlation for the calculation of intermolecular electrostatic interaction energies, (b) almost complete convergence of E_{es} at the MP4SDQ level, and (c) the relative unimportance of higher-order terms included in the CCSD theory.[63] We find that intramolecular correlation included even at the MP2 level yields highly satisfactory electrostatic interaction energies for the type of systems studied here.

3. Effectiveness of the Databank in the E_{es} Calculation. One of the goals of this study is to obtain reliable reference values for E_{es} in the test dimers in order to provide a benchmark of accuracy for E_{es} obtained with the DB+EPMM approach. Two questions have to be addressed: (1) how does the databank approach compare with the B3LYP/6-31G** method on which it is based, and (2) how does it compare with much more advanced correlated methods?

As to the first question, the agreement between electrostatic interaction energies calculated with the DB+EPMM method and B3LYP/6-31G** values is quite good −under 4 kJ/mol (~1 kcal/mol) for five out of six dimers. This good agreement should be viewed in light of the fact that the glycine molecule was not included in the set of molecules used in the construction of the pseudoatom databank. For the Gly3 dimer, the difference is slightly larger−9 kJ/mol.

Taking account of the fact that, in constructing the databank, B3LYP/6-31G** Gaussian-type densities were projected onto the Slater-type basis set used in the Hansen−Coppens pseudoatom model, and that the final set of pseudoatom parameters is obtained by averaging over many slightly different chemical environments and atomic conformations, a root-mean-square (RMS) discrepancy of 4 kJ/mol is quite acceptable.

As to the second question, DB+EPMM, like B3LYP/6-31G** itself, always underestimates the attractive electrostatic interaction energy compared to our best ab initio CCSD/aug-cc-pVTZ calculation. The differences can be fairly significant. Thus, for dimers Gly3 and Gly4, the differences between DB+EPMM and CCSD/aug-cc-pVTZ E_{es} values are as large as ∼30 and 20 kJ/mol, respectively. But, for the Gly1, Gly2, and Gly6 dimers, the DB+EPMM approach underestimates E_{es} by only 5−7 kJ/mol (1−2 kcal/mol). For the only repulsive dimer, Gly5, the DB+EPMM energy is in excellent agreement with the CCSD/aug-cc-pVTZ value. The RMS discrepancy between DB+EPMM and CCSD/aug-cc-pVTZ E_{es} values is only 16 kJ/mol for the set of six dimers, essentially due to the less advanced method on which the databank is based. For comparison, the RMS deviation between B3LYP/6-31G** and CCSD/aug-cc-pVTZ energies is 14 kJ/mol, and between the best ADF BLYP/QZ4P calculation and CCSD/aug-cc-pVTZ it is 9 kJ/mol.

4. Dependence of Electron Density Distributions on the Level of Theory. Electrostatic interaction energies described in this paper are, of course, intimately related to the electron density distribution in the monomer of α-glycine. Figure 5a shows the difference between HF/cc-pVTZ and HF/cc-pVDZ electron density distributions plotted in the plane of an oxygen, the carbon atom of the CH_2 group, and the nitrogen atom. The extension of the basis set from DZ to TZ significantly affects the spherical component of the electron density near the atom cores, which is expected to be relatively unimportant for E_{es} calculations, and increases the density in the bonding and tail regions of the density distributions, which is expected to be more important. Figure 5b illustrates the effect of including diffuse functions in the cc-pVDZ orbital basis set at the HF level. The results for other methods and basis sets are similar. Surprisingly, the effect is not confined to the tails of the density distributions but is also pronounced near the atoms and in the bonding regions. Most remarkable are the nonspherical features around the atoms. The effect of electron correlation is shown in Figure 5c. As observed in previous studies,[30] correlation builds charge density near the nuclei and decreases it in bonding regions. Contrary to earlier studies, the charge density is actually depleted in a very small region in the immediate vicinity of the oxygen and nitrogen atoms. To ensure that this feature is not an artifact of our calculations, we repeated the formaldehyde calculations previously reported by Wiberg et al.,[30] computing charge densities on a finer grid of points, and found the same feature in that molecule.

In general, the effects of the basis set and method of computation are rather significant and sufficiently complicated to account for the observed changes in the inter-

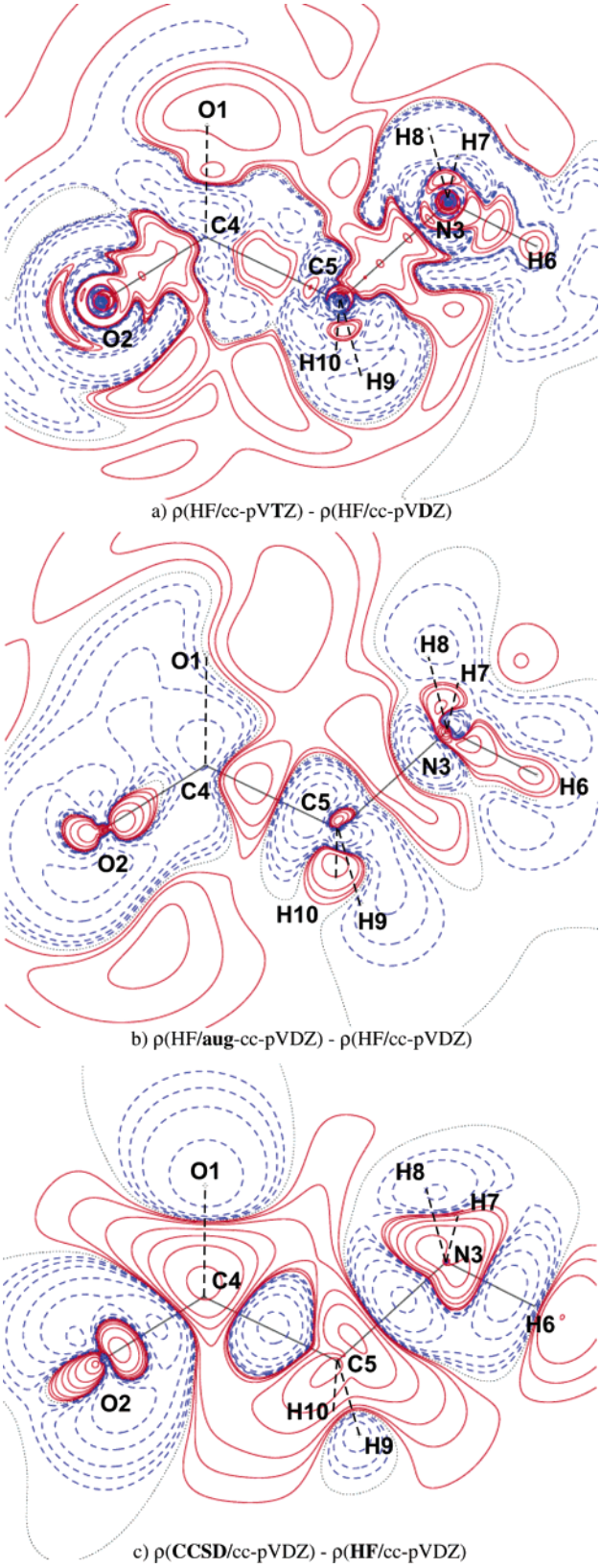


Figure 5. Differences in the charge density distribution in the glycine molecule between various levels of theory in the plane of the oxygen, the carbon of the CH_2 group, and the nitrogen atom. Positive contours are shown with a solid red line, negative with a dashed blue line, and zero with a dotted black line. Contour levels are $\pm 2 \times 10^{-4}$, $\pm 4 \times 10^{-4}$, $\pm 8 \times 10^{-4}$, $\pm 2 \times 10^{-3}$, $\pm 4 \times 10^{-3}$, $\pm 8 \times 10^{-3}$, $\pm 2 \times 10^{-2}$, and so forth e/au^3.

molecular electrostatic interaction energy and its orientational dependence.

Concluding Remarks

The results of an extensive analysis of the basis set and method dependence of intermolecular electrostatic interaction energies in six dimers of α-glycine show that the most significant effect on E_{es} in Gaussian-type calculations is produced by the inclusion of diffuse functions, even in the case of the fairly extended cc-pVQZ basis set. For augmented Gaussian basis sets, the convergence of E_{es} is nearly complete at the aug-cc-pVDZ level. The basis set dependence in Slater-type calculations of E_{es} is somewhat smaller than that for Gaussians.

The method dependence of the calculated E_{es} is also pronounced. Advanced correlated methods, such as QCISD, CCSD, and MP4SDQ, and also MP2, show very consistent results, usually within a range of 1−2 kJ/mol. Of these, MP2 is much less computationally demanding and when combined with aug-cc-pVTZ or (if possible) a larger basis set is capable of producing accurate benchmark electrostatic interaction energies. Electrostatic energies obtained with HF and CISD methods deviate considerably from these results, generally overestimating the magnitude of the electrostatic interaction. Pure DFT functionals with both Gaussian and Slater basis functions almost always show large deviations from advanced correlated methods, apparently because of an incorrect long-range behavior of these functionals. Despite their different origins, BLYP and PBE functionals yield very similar E_{es} energies. Electrostatic interaction energies from the hybrid DFT B3LYP functional are in much better agreement with those of advanced correlated methods, especially when augmented TZ- or QZ-type basis sets are used. This is due to the inclusion of Hartree−Fock exchange, which by itself contains correct asymptotics and improves the overall asymptotic behavior of a hybrid functional.

The combination of the pseudoatom databank and EPMM method is well able to reproduce the results of the B3LYP/6-31G** calculations on which it is based (well within 10 kJ/mol, usually within 4 kJ/mol only), confirming the transferability of the pseudoatom densities among the types of molecules considered. However, because electrostatic interaction energies calculated at the B3LYP/6-31G** level of theory deviate (sometimes by 20−30 kJ/mol) from advanced correlated results, the databank results show analogous discrepancies. This indicates that the databank can be improved by the use of electron densities from advanced correlated methods.

Nevertheless, the combination of the current databank for the evaluation of electrostatic interaction energies in molecular systems with high-quality atom−atom potentials for the description of exchange−repulsion, dispersion, and induction forces should provide total bonding energies at an accuracy similar to or better than those obtained by the standard DFT methods. This approach is now being pursued.[64]

Acknowledgment. We would like to thank Prof. Krzysztof Szalewicz (Physics Department, University of Delaware) for stimulating discussions, many valuable suggestions, and encouragement of our work and Dr. Matt Jones (Center for Computational Research, University at Buffalo) for his help with parallelizing the SPDFG program. Support by the National Institutes of Health (GM56829) and the National Science Foundation (CHE0236317) is gratefully acknowledged. Molecular graphics have been made with the program MOLEKEL.[65,66]

References

(1) Xu, D.; Lin, S. L.; Nussinov, R. *J. Mol. Biol.* **1997**, *265*, 68−84.

(2) Stone, A. J. *The Theory of Intermolecular Forces*; Oxford University Press: Oxford, England, 1996; International Series of Monographs on Chemistry 32.

(3) Jeziorski, B.; Moszynski, R.; Szalewicz, K. *Chem. Rev.* **1994**, *94*, 1887−1930.

(4) Bickelhaupt, F. M.; Baerends, E. J. *Rev. Comput. Chem.* **2000**, *15*, 1−86.

(5) Volkov, A.; Coppens, P. *J. Comput. Chem.* **2004**, *25*, 921−934.

(6) Buckingham, A. D. *Adv. Chem. Phys.* **1967**, *12*, 107−142.

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

(8) MacKerell, A. D.; Wiorkiewicz-Kuczera, J.; Karplus. M. *J. Am. Chem. Soc.* **1995**, *117*, 11946−11975.

(9) Halgren, T. A. *J. Comput. Chem.* **1996**, *17*, 490−519.

(10) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **2001**, *105*, 6474−6487.

(11) Allinger, N. L.; Yuh, Y. H.; Lii, J.-H. *J. Am. Chem. Soc.* **1989**, *111*, 8551−8566.

(12) Stone, A. J. *Chem. Phys. Lett.* **1981**, *83*, 233−239.

(13) Stone, A. J.; Alderton, M. *Mol. Phys.* **1985**, *56*, 1047−1064.

(14) Spackman, M. A. *J. Chem. Phys.* **1986**, *85*, 6587−6601.

(15) Volkov, A.; Koritsanszky, T. S.; Coppens, P. *Chem. Phys. Lett.* **2004**, *391*, 170−175.

(16) Hansen, N. K.; Coppens, P. *Acta Crystallogr., Sect. A* **1978**, *34*, 909−921.

(17) Coppens, P. *X-ray Charge Densities and Chemical Bonding*; Oxford University Press: New York, 1997.

(18) Koritsanszky, T.; Volkov, A.; Coppens, P. *Acta Crystallogr., Sect. A* **2002**, *58*, 464−472.

(19) Volkov, V.; Li, X.; Koritsanszky, T.; Coppens, P. *J. Phys. Chem. A* **2004**, *108*, 4283−4300.

(20) Ziegler, T.; Rauk, A. *Theor. Chim. Acta* **1977**, *46*, 1−10.

(21) Morokuma, K. *J. Chem. Phys.* **1971**, *55*, 1236−1244.

(22) te Velde, G.; Bickelhaupt, F. M.; van Gisbergen, S. J. A.; Fonseca Guerra, C.; Baerends, E. J.; Snijders, J. G.; Ziegler, T. *J. Comput. Chem.* **2001**, *22*, 931−967.

(23) Fonseca Guerra, C.; Snijders, J. G.; te Velde, G.; Baerends, E. *Theor. Chem. Acc.* **1998**, *99*, 391−403.

(24) *ADF*, version 2004.01; SCM, Theoretical Chemistry, Vrije Universiteit: Amsterdam, The Netherlands, 2004. http://www.scm.com.

(25) Destro, R.; Roversi, P.; Barzaghi, M.; Marsh, R. E. *J Phys Chem A* **2000**, *104*, 1047−1054.

(26) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; 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.01; Gaussian, Inc.: Wallingford, CT, 2004.

(27) Handy, N. C.; Schaefer, H. F., III. *J. Chem. Phys.* **1984**, *81*, 5031−5033.

(28) Diercksen, G. H. F.; Roos, B. O.; Sadlej, A. J. *Chem. Phys.* **1981**, *59*, 29−39.

(29) Diercksen, G. H. F.; Sadlej, A. J. *J. Chem. Phys.* **1981**, *75*, 1253−1266.

(30) Wiberg, K. B.; Hadad, C. M.; LePage, T. J.; Breneman, C. M.; Frisch, M. J. *J. Phys. Chem.* **1992**, *96*, 671−679.

(31) Dupuis, M.; Rys, J.; King, H. F. *J. Chem. Phys.* **1976**, *65*, 111−116.

(32) Rys, J.; Dupuis, M.; King, H. F. *J. Comput. Chem.* **1983**, *4*, 154−157.

(33) King, H. F.; Dupuis, M. *J. Comput. Phys.* **1976**, *21*, 144−165.

(34) Hariharan, P. C.; Pople, J. A. *Theor. Chim. Acta* **1973**, *28*, 213−222.

(35) Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257−2261.

(36) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098−3100.

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

(38) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1997**, *78*, 1396−1396.

(39) Dunning, T. H., Jr. *J. Chem. Phys.* **1970**, *53*, 2823−2833.

(40) Dunning, T. H., Jr.; Hay, P. J. In *Methods of Electronic Structure Theory*; Schaefer, H. F., III, Ed.; Plenum Press: New York, 1977; Vol. 3.

(41) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648−5652.

(42) Dunning, T. H., Jr. *J. Chem. Phys.* **1989**, *90*, 1007−1023.

(43) Davidson, E. R. *Chem. Phys. Lett.* **1996**, *260*, 514−518.

(44) Woon, D. E.; Dunning, T. H., Jr. *J. Chem. Phys.* **1994**, *100*, 2975−2988.

(45) Kendall, R. A.; Dunning, T. H., Jr.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796−6806.

(46) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347−1363.

(47) Rybak, S.; Jeziorski, B.; Szalewicz, K. *J. Chem. Phys.* **1991**, *95*, 6576−6601.

(48) Williams, H. L.; Mas, E. M.; Szalewicz, K.; Jeziorski, B. *J. Chem. Phys.* **1995**, *103*, 7374−7391.

(49) Szalewicz, K.; Cole, S. J.; Koos, W.; Barlett, R. J. *J. Chem. Phys.* **1988**, *89*, 3662−3673.

(50) Lii, J.-H.; Ma, B.; Allinger, N. L. *J. Comput. Chem.* **1999**, *20*, 1593−1603.

(51) Halkier, A.; Koch, H.; Jørgensen, P.; Christiansen, O.; Beck Nielsen, I. M.; Helgaker, T. *Theor. Chem. Acc.* **1997**, *97*, 150−157.

(52) Sinnokrot, M. O.; Valeev, E. F.; Sherrill, C. D. *J. Am. Chem. Soc.* **2002**, *124*, 10887−10893.

(53) van Duijneveldt-van de Rijdt, J. G. C. M.; van Duijneveldt, F. B. *J. Mol. Struct.* **1982**, *89*, 185−201.

(54) Sokalski, W. A.; Sawaryn, A. *J. Chem. Phys.* **1987**, *87*, 526−534.

(55) Moszynski, R.; Jeziorski, B.; Ratkiewicz, A.; Rybak, S. *J. Chem. Phys.* **1993**, *99*, 8856−8869.

(56) Amos, R. D. *Chem. Phys. Lett.* **1980**, *73*, 602−606.

(57) Misquitta, A. J.; Szalewicz, K. *Chem. Phys. Lett.* **2002**, *357*, 301−306.

(58) Misquitta, A. J.; Szalewicz, K. *J. Chem. Phys.* **2005**, *122*, 214109-1−214109-19.

(59) Tozer, D. J.; Handy, N. C. *J. Chem. Phys.* **1998**, *109*, 10180−10189.

(60) Casida, M. E.; Salahub, D. R. *J. Chem. Phys.* **2000**, *113*, 8918−8935.

(61) Szalewicz, K. Personal communication.

(62) Bukowski, R.; Szalewicz, K.; Chabalowski, C. F. *J. Phys. Chem.* **1999**, *103*, 7322−7340.

(63) Korona, T.; Moszynski, R.; Jeziorski, B. *Mol. Phys.* **2002**, *100*, 1723−1734.

(64) Volkov, A.; Li, X.; Coppens, P.; Szalewicz, K. To be published.

(65) Flükiger, P. F. Development of the molecular graphics package MOLEKEL and its application to selected problems in organic and organometallic chemistry, Thèse No 2561, Département de chimie physique, Université de Genève, Genève, Switzerland, 1992.

(66) Portmann, S.; Lüthi, H. P. *CHIMIA* **2000**, *54*, 766−769.

CT050216X

JCTC Journal of Chemical Theory and Computation

A Molecular Energy Decomposition Scheme for Atoms in Molecules

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

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

Received September 5, 2005

Abstract: An exact energy partition method based on a physically sound decomposition of the nondiagonal first-order and diagonal second-order density matrices put forward by Li and Parr (*J. Chem. Phys.* **1986**, *84*, 1704) is presented. The method splits the total energy into intra- and interatomic components and is applicable on quite general wave functions. To explore it numerically, the energy components of three test molecules (H_2, N_2, and LiH) have been computed using four different partitions of the charge density $\rho(\mathbf{r})$ into atomic densities. Several aspects on the chemical bond and the relative importance of different components of the binding energy are analyzed. The merits of different partitions of $\rho(\mathbf{r})$ are also discussed.

I. Introduction

Chemists usually see molecules as formed by atoms or groups of atoms interacting with each other in 3D space and approximately having transferable properties. This idea has inspired the translation of the well-known concepts of chemistry such as bonds, valences, atomic charges, and so forth to the quantum mechanical language. In looking for this connection, the proposed models usually start at a qualitative level, but as soon as they slide into the quantitative realm, they fall into a fragment picture. Many of these quantitative models try, then, energy partition schemes or energy decomposition analyses.[1−22]

All of these methods, which have importantly contributed to deepening our knowledge of the chemical bond, are not free from criticisms. A number of them are (a) their link to particular calculational procedures, (b) their dependence on the reference used to describe the fragments, (c) their use of fictitious intermediate states, and (d) their mixing of exchange and orthogonality constraints.

Recently, we presented an energy partition method that is theoretically sound and able to give detailed definitions of the interactions among atoms, functional groups, and molecules.[23] Moreover, it is exhaustive in the sense that it recovers exactly the total energy of the system and derives from the molecular wave function without resorting to the approximations involved in its calculation. Its implementation was possible thanks to the previous development of an efficient algorithm to compute two-electron integrals over arbitrary regions of space for both monodeterminantal[24] and correlated[25] wave functions. In ref 23, the atomic regions of the quantum theory of atoms in molecules (QTAM), mainly developed by Bader,[11] were taken as basic entities from which the molecule is built. The QTAM atoms, unequivocally defined as the 3D attraction basins of the gradient field of the molecular charge density $\rho(\mathbf{r})$, have sharp and well-defined boundaries and, thus, produce noninterpenetrating atomic densities. Their irregular forms and the high computational cost which is necessary to determine their boundaries has prevented their wide use in routine quantum chemical applications. For this reason, the objective of this paper is to present a molecular energy decomposition scheme that, taking also the individual atoms as the chemically meaningful fragments in the molecule, generalizes the earlier one based on QTAM atoms.[23] The basic idea of the generalized approach is to partition $\rho(\mathbf{r})$ in terms of interpenetrating atomic densities $\rho_A(\mathbf{r})$ that have no defined boundaries. The charge density at any point of physical space is not assigned to a single atom as in QTAM but is shared to a certain degree by all the atoms of the molecule. Since each $\rho_A(\mathbf{r})$ extends to infinity, the obstacles associated with the steplike character of QTAM atoms clearly disappear and

* Corresponding author. Phone: +34 985103039. E-mail: evelio@carbono.quimica.uniovi.es.

10.1021/ct0502209 CCC: $33.50
Published on Web 12/02/2005

the task of determining atomic boundaries is obviously absent. The only requisite that the $\rho_A(\mathbf{r})$ has to satisfy is

$$\rho(\mathbf{r}) = \sum_A w_A(\mathbf{r})\,\rho(\mathbf{r}) = \sum_A \rho_A(\mathbf{r}) \tag{1}$$

where the $w_A(\mathbf{r})$ functions provide a partition of the unity

$$\sum_A w_A(\mathbf{r}) = 1\ \forall \mathbf{r} \tag{2}$$

A partition of $\rho(\mathbf{r})$ into atomic components does not unequivocally define a corresponding energy partition since the total energy depends not only on $\rho(\mathbf{r})$ but also on the nondiagonal part of the first-order density matrix, $\rho_1(\mathbf{r},\mathbf{r}')$, and the diagonal second-order density matrix, $\rho_2(\mathbf{r}_1,\mathbf{r}_2)$. Consequently, a second and essential step in the energetic partition that we propose is to convey the partition of $\rho(\mathbf{r})$ to the full nondiagonal density matrices from which the total energy depends. This should be done, in our opinion, using physically reasonable arguments. We will use, in this article, the scheme proposed in the 1980s by Li and Parr.[26] It is very relevant to remark that Li and Parr's scheme allows for a physically sound partition of $\rho_1(\mathbf{r},\mathbf{r}')$ and $\rho_2(\mathbf{r}_1,\mathbf{r}_2)$ only in terms of a given partition of $\rho(\mathbf{r})$.

The third and last step in our partition method is to group or reorganize the different energy components into physically and chemically meaningful contributions. Altogether, these three elements represent a practical and physically well-founded methodology to partition the total energy, energy components, and density matrices into intra-atomic and interatomic terms. As we will see below, the algorithm works equally well with very different and unconnected partitions of $\rho(\mathbf{r})$: interpenetrating (in both its localized and delocalized versions) and noninterpenetrating (in which the atomic densities are exactly zero outside a given 3D region).

We have organized the rest of the paper as follows. In Section II, we present the energy partition method. In Section III, we present and discuss the results. First, taking the CO molecule as a test example, we analyze its atomic charges and densities and discuss the results found for these properties in other molecules (Subsection IIIA). Then, we perform a thorough comparative study of the energy components in H_2 using four possible partitions of $\rho(\mathbf{r})$ and show how their relative values are governed by the hydrogen atomic density in each of the partitions (Subsection IIIB). Similar studies on the N_2 and LiH molecules (representative of the traditional apolar covalent and partially ionic bonding types, respectively) are presented in Subsections IIIC and IIID, respectively. Finally, a summary and our conclusions are given in Section IV.

II. Energy and Charge Density Partitions

In this section, we present some theoretical aspects of the energy partition method that we propose (Subsection IIA), show how the different energy components can be rearranged to obtain deeper insights into their physical meaning (Subsection IIB), define the partitions of the charge density into atomic densities with which the energy partition has actually been applied (Subsection IIC), and give some relevant computational details concerning the practical evaluation of all the energy components (Subsection IID).

A. Energy Partition. Since the nonrelativistic Born−Oppenheimer molecular Hamiltonian contains only one- and two-particle terms, the total energy of a molecule may be obtained from just the spin-free first order, $\rho_1(\mathbf{r},\mathbf{r}')$, and diagonal second order, $\rho_2(\mathbf{r}_1,\mathbf{r}_2)$, reduced density matrices as[27]

$$E = E_e + V_{nn} = (T + V_{ne} + V_{ee}) + V_{nn} \tag{3}$$

where

$$T = \int_{\mathbf{r}'\to\mathbf{r}} \hat{T}\rho_1(\mathbf{r},\mathbf{r}')\,d\mathbf{r} \tag{4}$$

$$V_{ne} = -\sum_A Z_A \int \frac{\rho(\mathbf{r})}{|\mathbf{r}-\mathbf{R}_A|}\,d\mathbf{r} \tag{5}$$

$$V_{ee} = \frac{1}{2}\int \frac{\rho_2(\mathbf{r}_1,\mathbf{r}_2)}{r_{12}}\,d\mathbf{r}_1\,d\mathbf{r}_2 \tag{6}$$

and

$$V_{nn} = \frac{1}{2}\sum_A\sum_{B\neq A} V_{nn}^{AB} = \frac{1}{2}\sum_A\sum_{B\neq A}\frac{Z_A Z_B}{R_{AB}} \tag{7}$$

are the total kinetic energy, nucleus−electron attractive potential energy, electron−electron repulsion energy, and nuclei−nuclei repulsion energy, respectively, $\hat{T} = {}^1/_2\nabla\cdot\nabla'$, and $\rho(\mathbf{r}) \equiv \rho_1(\mathbf{r},\mathbf{r})$. What we want is a consistent partition of $\rho_1(\mathbf{r},\mathbf{r}')$ and $\rho_2(\mathbf{r}_1,\mathbf{r}_2)$ (and from it, a partition of the total energy) using exclusively a well-defined partition of $\rho(\mathbf{r})$ into atomic densities $\rho_A(\mathbf{r})$'s. A method to do this was proposed 20 years ago by Li and Parr.[26] Following these authors, we assume the validity of eq 1 also for the nondiagonal $\rho_1(\mathbf{r},\mathbf{r}')$ and partition it in the form

$$\rho_1(\mathbf{r},\mathbf{r}') = \sum_A w_A(\mathbf{r}')\,\rho_1(\mathbf{r},\mathbf{r}') = \sum_A \rho_1^A(\mathbf{r},\mathbf{r}') \tag{8}$$

where the $w_A(\mathbf{r})$'s satisfy eq 2. It may then be shown that

$$\left.\frac{\hat{T}\rho_1^A(\mathbf{r},\mathbf{r}')}{\rho_1^A(\mathbf{r},\mathbf{r}')}\right|_{\mathbf{r}'\to\mathbf{r}} = \left.\frac{\hat{T}\rho_1(\mathbf{r},\mathbf{r}')}{\rho_1(\mathbf{r},\mathbf{r}')}\right|_{\mathbf{r}'\to\mathbf{r}} \quad \forall A \tag{9}$$

In Li and Parr's original scheme, based on density functional theory (DFT), this amounts, to scale, the exact kinetic density functional. A very similar scaling is done to partition $\rho_2(\mathbf{r}_1,\mathbf{r}_2)$, this time, with a double scaling for electrons 1 and 2:

$$\rho_2(\mathbf{r}_1,\mathbf{r}_2) = \sum_A\sum_B w_A(\mathbf{r}_1)\,w_B(\mathbf{r}_2)\,\rho_2(\mathbf{r}_1,\mathbf{r}_2) \tag{10}$$

$$= \sum_A\sum_B \rho_2^{AB}(\mathbf{r}_1,\mathbf{r}_2) \tag{11}$$

This partition implies that

$$\frac{\rho_2^{AA}(\mathbf{r}_1,\mathbf{r}_2)}{\rho_A(\mathbf{r}_1)\,\rho_A(\mathbf{r}_2)} = \frac{\rho_2^{AB}(\mathbf{r}_1,\mathbf{r}_2)}{\rho_A(\mathbf{r}_1)\,\rho_B(\mathbf{r}_2)} = \frac{\rho_2(\mathbf{r}_1,\mathbf{r}_2)}{\rho(\mathbf{r}_1)\,\rho(\mathbf{r}_2)} \tag{12}$$

which means that, given two arbitrary points $\mathbf{r}_1$ and $\mathbf{r}_2$, the interaction energy between two electrons is the same no matter whether we assume them as belonging to an atom or to the molecule.

These ideas lead to a partition of all the energy components into intra- and interatomic contributions. For instance, using eqs 2 and 8 in eq 4, we obtain

$$T = \sum_A T^A = \sum_A \int_{\mathbf{r}'\to\mathbf{r}} w_A(\mathbf{r})\, \hat{T}\rho_1(\mathbf{r},\mathbf{r}')\, d\mathbf{r} \tag{13}$$

Similarly, when eq 1 is used in eq 5, V_{ne} is given as

$$V_{ne} = -\sum_A \sum_B Z_A \int \frac{\rho_B(\mathbf{r})}{|\mathbf{r}-\mathbf{R}_A|}\, d\mathbf{r} = \sum_A \sum_B V_{ne}^{AB} \tag{14}$$

Finally, inserting eq 11 in eq 6, V_{ee} results:

$$V_{ee} = \sum_A V_{ee}^{AA} + \frac{1}{2}\sum_A \sum_{B\neq A} V_{ee}^{AB} \tag{15}$$

where

$$V_{ee}^{AA} = \frac{1}{2}\int \frac{\rho_2^{AA}(\mathbf{r}_1,\mathbf{r}_2)}{r_{12}}\, d\mathbf{r}_1\, d\mathbf{r}_2 \tag{16}$$

and

$$V_{ee}^{AB} = \int \frac{\rho_2^{AB}(\mathbf{r}_1,\mathbf{r}_2)}{r_{12}}\, d\mathbf{r}_1\, d\mathbf{r}_2 \tag{17}$$

Inserting now eqs 7, 13, 14, and 15 in eq 3, the total energy can be expressed as

$$E = \sum_A E_{net}^A + \frac{1}{2}\sum_A \sum_{B\neq A} E_{int}^{AB} \tag{18}$$

where

$$E_{net}^A = T^A + V_{ne}^{AA} + V_{ee}^{AA} \tag{19}$$

is the net energy of atom A and

$$E_{int}^{AB} = V_{ne}^{AB} + V_{ne}^{BA} + V_{ee}^{AB} + V_{nn}^{AB} \tag{20}$$

is the total interaction energy between atoms A and B. Each atomic net energy, E_{net}^A, is an effective one-body term that carries all the intra-atomic energy contributions, while each interaction term, E_{int}^{AB}, is an effective two-body component of the total energy. Both contributions actually include all the many-body interactions that result from a quantum-mechanical calculation. However, since the molecular Hamiltonian is expressed as a sum of one- and two-particle terms only, the total energy does not contain explicit many-body interactions.

Equation 18 defines the present energy partition. It states that the total energy can be exactly written as a sum of the net energies of all the atoms of the system and the interatomic interaction energies. It has been derived from the molecular wave function without resorting to the approximations involved in its calculation or to the specific peculiarities of the orbital description used to obtain it. Moreover, the partition is general in the sense that it can be applied with any definition of the atomic densities $\rho^A(\mathbf{r})$, provided that they satisfy eq 1

B. Analysis of the Energy Components. Both E_{net}^A (eq 19) and E_{int}^{AB} (eq 20) can be partitioned into more detailed components with a clear physical meaning. This is possible thanks to the natural decomposition of $\rho_2(\mathbf{r}_1,\mathbf{r}_2)$ in Coulomb and exchange-correlation components:

$$\rho_2(\mathbf{r}_1,\mathbf{r}_2) = \rho_2^C(\mathbf{r}_1,\mathbf{r}_2) + \rho_2^{xc}(\mathbf{r}_1,\mathbf{r}_2) \tag{21}$$

where $\rho_2^C(\mathbf{r}_1,\mathbf{r}_2) = \rho(\mathbf{r}_1)\,\rho(\mathbf{r}_2)$. On the other hand, even though the meaning of exchange and correlation energies is strictly lost as soon as correlated wave functions are used,[28] a reasonable separation of both terms can be performed, defining the exchange density matrix, $\rho_2^X(\mathbf{r}_1,\mathbf{r}_2)$, as in a monodeterminantal case (Fock−Dirac exchange)

$$\rho_2^X(\mathbf{r}_1,\mathbf{r}_2) = -\rho_1(\mathbf{r}_2,\mathbf{r}_1)\,\rho_1(\mathbf{r}_1,\mathbf{r}_2) \tag{22}$$

and the correlation density matrix, $\rho_2^{corr}(\mathbf{r}_1,\mathbf{r}_2)$, by difference

$$\rho_2^{corr}(\mathbf{r}_1,\mathbf{r}_2) = \rho_2 - \rho_2^C - \rho_2^X = \rho_2^{xc} - \rho_2^X \tag{23}$$

Equations 21−23 allow us to write V_{ee}^{AB} as

$$V_{ee}^{AB} = V_C^{AB} + V_{xc}^{AB} \tag{24}$$

$$V_{ee}^{AB} = V_C^{AB} + V_X^{AB} + V_{corr}^{AB} \tag{25}$$

where

$$V_\tau^{AB} = \int\int \frac{\rho_2^{\tau,AB}(\mathbf{r}_1,\mathbf{r}_2)}{r_{12}}\, d\mathbf{r}_1\, d\mathbf{r}_2 \tag{26}$$

$\tau = C$, xc, X, or corr, and

$$\rho_2^{\tau,AB}(\mathbf{r}_1,\mathbf{r}_2) = \rho_2^\tau(\mathbf{r}_1,\mathbf{r}_2)\, w_A(\mathbf{r}_1)\, w_B(\mathbf{r}_2) \tag{27}$$

When eqs 21−23 are used, the intra-atomic repulsion energy V_{ee}^{AA} (eq 16) may also be expressed as a sum of Coulomb, exchange, and correlation contributions:

$$V_{ee}^{AA} = V_C^{AA} + V_{xc}^{AA} \tag{28}$$

$$V_{ee}^{AA} = V_C^{AA} + V_X^{AA} + V_{corr}^{AA} \tag{29}$$

where the expression for V_τ^{AA} is similar to that of V_τ^{AB} (eq 26) but using $^1/_2\rho_2^{AA}$ instead of ρ_2^{AB}.

The atomic net energies (eq 19) contain the same energetic contributions present in the isolated atoms. When the atomic density does not change much for a given atom in different molecules, its energy components (kinetic energies, nuclear attraction to its atomic nucleus, and electron−electron repulsion) will be largely the same. Consequently, the net energies carry the atomic identity from system to system. The deformation energy

$$E_{def}^A = E_{net}^A - E_0^A \tag{30}$$

where E_0^A is the net energy of atom A in vacuo, provides a measurement of the change suffered by an atom in going from the isolated state to the molecule. It plays an important role in the definition of the molecular binding energy, E_{bind}. This property is defined as the total molecular energy referred to an appropriate reference. Taking the isolated neutral atoms' reference, we have

$$E_{\text{bind}} = E - \sum_A E_0^A \tag{31}$$

and using eqs 18 and 30, we obtain

$$E_{\text{bind}} = \sum_A E_{\text{def}}^A + \frac{1}{2}\sum_A \sum_{B\neq A} E_{\text{int}}^{AB} \tag{32}$$

The stability of a molecule with respect to its atomic components is, thus, determined by two factors: the deformation energy, which is necessarily positive in homonuclear diatomic molecules[26] and is usually positive (provided that the neutral atoms are taken as references to define E_{bind}) in many other cases, and the interaction between the atoms, which is usually negative when *A* and *B* are bonded.

When eqs 24 and 25 are used in eq 20, the total interaction energy, E_{int}^{AB}, can be written as

$$E_{\text{int}}^{AB} = V_{\text{cl}}^{AB} + V_{\text{xc}}^{AB} \tag{33}$$

$$E_{\text{int}}^{AB} = V_{\text{cl}}^{AB} + V_X^{AB} + V_{\text{corr}}^{AB} \tag{34}$$

where

$$V_{\text{cl}}^{AB} = V_{\text{ne}}^{AB} + V_{\text{ne}}^{BA} + V_{\text{nn}}^{AB} + V_C^{AB} \tag{35}$$

is the classical electrostatic Coulomb interaction and V_{xc}^{AB} is the interaction energy due to purely quantum effects (i.e., exchange and correlation). This rearrangement is very convenient since the four components of E_{int}^{AB} in eq 20 are usually orders of magnitude larger than the interaction energy itself. However, V_{cl}^{AB}, which can be written in the compact form

$$V_{\text{cl}}^{AB} = \int \frac{\rho_A^T(\mathbf{r}_1)\,\rho_B^T(\mathbf{r}_2)}{r_{12}}\,d\mathbf{r}_1\,d\mathbf{r}_2 \tag{36}$$

where $\rho_A^T(\mathbf{r}) = Z_A\delta(\mathbf{r} - \mathbf{R}_A) - \rho_A(\mathbf{r})$ is the total (nuclear plus electron) charge density of atom *A*, will always be much smaller than each of the individual terms in eq 35.[23] In fact, it can be shown that V_{cl}^{AB} is necessarily positive for two non-interpenetrating and specular distributions of charge. As we will see later, this means that $V_{\text{cl}}^{AB} > 0$ for the two QTAM atoms of a homonuclear diatomic molecule. Since, in these molecules, $E_{\text{def}}^A > 0$, we conclude that the stability of a homonuclear diatomic molecule in the QTAM energy partitioning scheme is a pure quantum phenomenon; that is, it is exclusively due to the interatomic exchange-correlation stabilizing interactions, the rest of energetic components being overall repulsive. We will also see below that this is not necessarily true when overlapping atomic densities $\rho_A(\mathbf{r})$ are used to construct $\rho(\mathbf{r})$.

A measure of the delocalization of electrons of atom *A* in atom *B* and vice versa is given by

$$F_{AB} = 2\int \rho_2^{\text{xc},AB}(\mathbf{r}_1,\mathbf{r}_2)\,d\mathbf{r}_1\,d\mathbf{r}_2 \tag{37}$$

For each *AB* pair, $\delta_{AB} = |F_{AB}|$ is the delocalization index, which can be roughly interpreted as the number of electron bonding pairs shared by the atoms. In this sense, δ_{AB} is a good quantum-mechanical indicator of covalency.

C. Charge Density Partitions. There exists an arbitrary number of ways to partition $\rho(\mathbf{r})$ into atomic densities.[11,26,29,30] One of the methods more firmly rooted in the basic principles of quantum mechanics is the exhaustive partition of $\mathbb{R}^3$ into proper open quantum systems provided by the QTAM of Bader and co-workers.[11,31,32] The theory divides the space into the 3D attraction basins of the gradient field of $\rho(\mathbf{r})$. These atomic basins Ω_A usually contain one and only one nucleus, and they are bounded by a zero local flux surface of $\nabla\rho$ [$\nabla\rho(\mathbf{r})\cdot\mathbf{n}(\mathbf{r}) = 0$ for $\mathbf{r} \in S(\Omega_A)$, where $\mathbf{n}(\mathbf{r})$ is a vector normal to the surface $S(\Omega_A)$]. The QTAM partition can be recast in the form given in eq 1 by simply choosing $w_A(\mathbf{r}) = 1$ for $\mathbf{r} \in \Omega_A$ and $w_A(\mathbf{r}) = 0$ for $\mathbf{r} \notin \Omega_A$.

The QTAM atoms have sharp and well-defined boundaries, present in some cases in irregular forms, and are computationally very costly to determine. Some of these inconveniences can be avoided in the partitions of $\rho(\mathbf{r})$ on the basis of interpenetrating atomic densities (IADs).[26,29,30]

Among the many possibilities, the IADs proposed by Hirshfeld[29] are those preserving as much as possible of the information contained in the charge densities of the isolated atoms.[33] In this partition, each $w_A(\mathbf{r})$ is defined as the ratio of the in vacuo charge density of atom *A* to that of the *promolecule* (set of in vacuo atoms placed at the positions of the nuclei in the actual molecule), that is, $w_A(\mathbf{r}) = \rho_A^0(\mathbf{r})/\sum_A \rho_A^0(\mathbf{r})$. Since the atomic densities of the isolated atoms are required to determine the $w_A(\mathbf{r})$ values, neither an energy partition nor a population analysis can be performed within this scheme based only on the molecular wave function. One is necessarily forced to choose some external atomic densities. To avoid this requirement, we have considered, in this work, the following variation of the Hirshfeld's partition. Taking into account that $\rho(\mathbf{r})$ is usually given in terms of one-center and two-center contributions

$$\rho(\mathbf{r}) = \sum_A \rho_A^0(\mathbf{r}) + \sum_{A\neq B} \rho_{AB}^0(\mathbf{r}) \tag{38}$$

$$\rho_A^0(\mathbf{r}) = \sum_a \sum_{a'} \rho_{aa'}^{AA}\,\phi_a^A(\mathbf{r})\,\phi_{a'}^A(\mathbf{r}) \tag{39}$$

$$\rho_{AB}^0(\mathbf{r}) = \sum_a \sum_b \rho_{ab}^{AB}\,\phi_a^A(\mathbf{r})\,\phi_b^B(\mathbf{r}) \tag{40}$$

where $\phi_a^A(\mathbf{r})$ is a primitive Gaussian function centered at nucleus *A*; we define $w_A(\mathbf{r}) = \tilde{\rho}_A^0(\mathbf{r})/\sum_A \tilde{\rho}_A^0(\mathbf{r})$, where $\tilde{\rho}_A^0(\mathbf{r}) = c_A\,\rho_A^0(\mathbf{r})$ and c_A is a constant chosen such that $\int \tilde{\rho}_A^0(\mathbf{r})\,d\mathbf{r} = Z_A$. Although this partition (Mod-H from now on) is formally equivalent to that of Hirshfeld, it only requires the wave function of the system. We want to remark, at this point, that the modified atomic density $\tilde{\rho}_A^0(\mathbf{r})$ is used

exclusively to define the $w_A(\mathbf{r})$, and they enter the partition into $\rho_A(\mathbf{r})$ only through $w_A(\mathbf{r})$, not through $\rho(\mathbf{r})$ (see eq 1).

Another widely used partition of $\rho(\mathbf{r})$ in terms of IADs is that of Becke,[34] which was initially proposed to simplify the numerical evaluation of monoelectronic multicenter integrals in DFT.[34] It consists of dividing the $\mathbb{R}^3$ space into atomic regions that resemble fuzzy Voronoi polyhedra. The size of each atomic region is adjusted by using an effective radius R_A for each of the atoms of the molecule. In the original work,[34] and also in most of the works that use this partition, R_A is taken as the Bragg−Slater radius of the isolated atom. However, this choice produces very unrealistic atomic charges in many cases. For this reason, we will also use here atomic radii derived from a topological analysis of $\rho(\mathbf{r})$. Provided that a bond critical point (BCP) of $\rho(\mathbf{r})$ exists between atoms A and B, R_A (R_B) is taken as the distance from atom A (B) to the BCP. The number of topological radii of a given atom is, thus, equal to the number of bonds of this atom. In the case that no BCP exists between atoms A and B, R_A is taken as in the original work, that is, as the Bragg−Slater radius of the isolated atom. Becke's partition of $\rho(\mathbf{r})$ based on topological radii is labeled B-Top in this work.

Finally, in the partition method recently developed by Rico et al.,[30] $\rho_A(\mathbf{r})$ is determined following a minimal deformation criterion (MinDef in what follows) for every two-center contribution to $\rho(\mathbf{r})$. Writing each of these contributions as $\rho_{ab}^{AB}\,\phi_a^A(\mathbf{r})\,\phi_b^B(\mathbf{r})$, where ρ_{ab}^{AB} is a density matrix coefficient and $\phi_a^A(\mathbf{r})$ and $\phi_b^B(\mathbf{r})$ are primitive Gaussian functions centered at A and B, respectively (with orbital exponents ζ_A and ζ_B), the MinDef method assigns its entire value to atom A if $\zeta_A > \zeta_B$ or to atom B if $\zeta_B > \zeta_A$. If both orbital exponents are equal, half of each two-center contribution is assigned to each center. In practical terms, the classical Mulliken's partition only differs from the MinDef method in that the former always performs a symmetric partition of each two-center contribution regardless of the values of ζ_A and ζ_B.

D. Computational Aspects. We have shown in ref 25 how, in many of the actual quantum mechanical molecular computations, ρ_2 and ρ_2^X can be written in the forms

$$\rho_2(\mathbf{r}_1,\mathbf{r}_2) = \sum_i^M \lambda_i\, F_i(\mathbf{r}_1)\, F_i(\mathbf{r}_2) \tag{41}$$

$$\rho_2^X(\mathbf{r}_1,\mathbf{r}_2) = \sum_i^M \eta_i\, G_i(\mathbf{r}_1)\, G_i(\mathbf{r}_2) \tag{42}$$

where $M = m(m+1)/2$, m is the number of partially or fully occupied (real) molecular orbitals ϕ_p in the wave function, each $F_i(\mathbf{r})$ and $G_i(\mathbf{r})$ is a known linear combination of products $\phi_p(\mathbf{r})\,\phi_q(\mathbf{r})$, and λ_i and η_i are also known coefficients.

The use of eqs 41 and 42 greatly reduces the computational effort of the two-electron integrations which are necessary to apply our energy partition method. All of these two-electron integrals over arbitrary regions of space can be efficiently computed for both monodeterminantal[24] and correlated[25] wave functions by means of an always-convergent generalization of the conventional multipolar approach (even for overlapping densities). In refs 24 and 25, the procedure was particularized to the QTAM atomic basins. However, the method can be applied as well to general tridimensional regions. In particular, it can be used with the fuzzy-boundary regions defined in the previous two subsections. The required computational effort can also be substantially reduced by computing and storing for further use the radial factors

$$R_{lm}^{\Omega_A}(r) = \left(\frac{4\pi}{2l+1}\right)^{1/2} \int_{\hat{r}} S_{lm}(\hat{r})\, f^{\Omega_A}(\mathbf{r})\, \mathrm{d}\hat{r} \tag{43}$$

for all the grid points of an appropriate radial quadrature. In eq 43, $\hat{r} = (\theta,\phi)$; $S_{lm}(\hat{r})$ is a real spherical harmonic, defined as in ref 24; and $f^{\Omega_A}(\mathbf{r})$ is $\rho_A(\mathbf{r})$, $F_i^A(\mathbf{r}) = w_A(\mathbf{r})\,F_i(\mathbf{r})$, or $G_i^A(\mathbf{r}) = w_A(\mathbf{r})\,G_i(\mathbf{r})$, where $F_i(\mathbf{r})$ [$G_i(\mathbf{r})$] is one the functions of eq 41 (eq 42). Let us recall that the bipolar expansion for r_{12}^{-1} used in this work is always convergent.[24] Nevertheless, simpler integration methods based on a standard multipolar expansion of r_{12}^{-1}, used for instance by Popelier et al.,[35−37] converge to the exact results for sufficiently separated atoms. In this standard multipolar approach, the Coulombic interaction between $\rho_A(\mathbf{r}_1)$ and $\rho_B(\mathbf{r}_2)$ is given by[24]

$$V_{\mathrm{C,lr}}^{AB} = \sum_{l_1m_1l_2m_2}^{\infty}\sum^{\infty} C_{l_1m_1l_2m_2} \frac{Q_{l_1m_1}^{\Omega_A} Q_{l_2m_2}^{\Omega_B}}{R^{l_1+l_2+1}} \tag{44}$$

where $C_{l_1m_1l_2m_2}$ is a coupling coefficient[24] and $Q_{lm}^{\Omega_A}$ are spherical atomic multipoles defined as

$$Q_{lm}^{\Omega A} = \left(\frac{4\pi}{2l+1}\right)^{1/2} \int r^l\, S_{lm}(\hat{r})\, \rho_A(\mathbf{r})\, \mathrm{d}\mathbf{r} \tag{45}$$

The differences between the approximate and exact multipoles or between V_{C}^{AB} (eq 26, exact) and $V_{\mathrm{C,lr}}^{AB}$ (eq 44, approximate) will indicate how strongly atoms A and B overlap.

III. Results and Discussion

In this section, we present the results of our energy partition method. First, we analyze the atomic densities and charges of both atoms of CO and comment on the results obtained in some other molecules (Subsection IIIA). In Subsection IIIB, we give a thorough analysis of the dihydrogen molecule, a paradigm in which any new idea or method should be tested on. Finally, the N_2 and LiH molecules, which may be considered representative of the covalent (N_2) and ionic (LiH) bonding types, will be analyzed.

All the calculations have used the *gamess* code[38] to obtain the wave functions and our *promolden* code to perform the energy partition. The wave functions have been computed in the ground electronic states using complete active space CAS[n,m] (n active electrons, m active orbitals) multiconfiguration calculations for H_2 ($CAS^{2,2}$), N_2 ($CAS^{10,10}$), and LiH ($CAS^{2,2}$) and a Hartree−Fock (HF) calculation for CO. Basis sets 6-311G(p), TZV(d), 6-311G(p), and TZV(2p,-3d)++ were used for H_2, N_2, LiH, and CO, respectively. The energy components in *promolden* have been computed to an accuracy of about 1.0−3.0 kcal/mol.

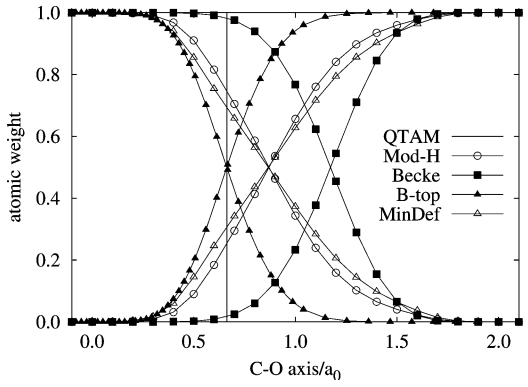


Figure 1. HF/TZV(2p,3d)++ atomic functions $w_A(\mathbf{r})$ for the carbon and oxygen atoms in the CO molecule along the internuclear axis. The C and O atoms are at the −0.0424 and 2.0424 positions along the C−O axis, respectively. Labels Becke, B-Top, MinDef, and Mod-H stand for the Becke with Bragg−Slater radii, Becke with topological radii, Rico et al. minimal deformation criterion, and modified Hirshfeld methods, respectively.

A. Atomic Overlapping Densities and Charges. The differences and similarities between the various partitions of $\rho(\mathbf{r})$ can be appreciated in Figure 1, where we have plotted $w_A(\mathbf{r})$ for both atoms in the CO molecule along the internuclear axis. In this figure, the QTAM $w_A(\mathbf{r})$ function is given by a vertical line at the BCP ($r_C = 0.705\ a_0$ and $r_O = 1.380\ a_0$). All the partitions generate localized atomic densities, and they basically differ in the size extension assigned to each atom. For a diatomic molecule, it is easy to show that, in the Becke and B-Top partitions, the point where $w_A = w_B = {}^1/_2$ along the $A-B$ axis satisfies $r_A/r_B = R_A/R_B$ [at this point $\rho_A(\mathbf{r}) = \rho_B(\mathbf{r}) = \rho(\mathbf{r})/2$, i.e., half of the charge density is assigned to atom A and half to atom B]. Consequently, when the Slater−Bragg radii of the isolated C and O atoms are used (0.65 a_0 and 0.47 a_0, respectively), this point is much closer to the oxygen atom ($r_O = 0.875\ a_0$) than to the carbon atom ($r_C = 1.210\ a_0$). However, when topological radii ($R_C^{top} = 0.705\ a_0$ and $R_O^{top} = 1.380\ a_0$) are used instead, that point moves to the BCP of the molecule. These numbers show that, in going from C to O along the C−O axis, $w_C(\mathbf{r})$ decays to zero much earlier in the B-Top than in the Becke partition, which means that a great quantity of electronic charge, ascribed to C in the Becke partition scheme, actually belongs to O in the B-Top method. This produces a change in the charge-transfer direction of the C−O bond in both cases ($C^{\delta-}O^{\delta+}$ in the Becke partition versus $C^{\delta+}O^{\delta-}$ in the B-Top partition).

In heterodiatomic molecules, the points in $\mathbb{R}^3$ for which $w_A = w_B = {}^1/_2$ in the B-Top partition do not necessarily lie on the QTAM interatomic surface. Of course, the BCP, which lies on this surface, is a notable exception since it satisfies the above property. Consequently, we expect that out of, but not very far from, the internuclear axis, the points in $\mathbb{R}^3$ for which $w_A = w_B = {}^1/_2$ will probably be relatively close to the QTAM interatomic surface. This explains the similarities between QTAM and B-Top atomic charges that we have found in many molecules. Of course, for homonuclear diatomic molecules, the points where $w_A = w_B = {}^1/_2$ always lie on the QTAM interatomic surface. Since, in the limit $k \to \infty$ (where k is the characteristic iterative parameter in the Becke partition method, see ref 34), $w_A(\mathbf{r})$ transforms to a steplike function [$w_A(\mathbf{r}) = 1$ for $r_A \le r_B$ and $w_A(\mathbf{r}) = 0$ for $r_A > r_B$], the QTAM and B-Top atoms, and all their properties and intra-atomic and interatomic interactions, tend to be equal as k increases in homonuclear diatomic molecules. This need not be so in heterodiatomics.

Table 1. HF/TZV(2p,3d)++ Atomic Charges for the Carbon and Oxygen Atoms of the CO Molecule from Different Partition Methods of $\rho(\mathbf{r})$[a]

	M	H	B	QTAM	Mod-H	B-Top	MinDef
C	0.448	0.124	−0.406	1.353	0.497	1.204	0.303
O	−0.448	−0.124	0.406	−1.356	−0.497	−1.204	−0.303

[a] M, H, and B letters stand for Mulliken, Hirshfeld, and Becke partitions, respectively. Labels B-Top, MinDef, Mod-H, and QTAM have been defined in the text.

The atomic charges derived by integrating $\rho_A(\mathbf{r})$ for the carbon and oxygen atoms of the CO molecule using the Becke, B-Top, MinDef, Mod-H, and QTAM partitions, as well as those obtained from the classical Mulliken and the original Hirshfeld partitions, are collected in Table 1. The Hirshfeld $w_A(\mathbf{r})$ functions were computed from the high-quality Koga's atomic densities of the isolated carbon and oxygen atoms.[39] The charge transfer (CT) predicted by the Becke partition ($C^{\delta-}O^{\delta+}$) is contrary to that obtained in all the other methods ($C^{\delta+}O^{\delta-}$). Analyzing the atomic charges obtained for many other molecules, we have observed that this partition behaves generally very differently from the rest, predicting, in many cases, a charge transfer that is even contrary to traditional chemical thinking. When a molecule is formed from neutral atoms, the effective atomic radii change with respect to their in vacuo values, this change increasing with the difference of electronegativities of both atoms. It seems that Becke's method does not properly account either for this changeor for the actual charge-transfer phenomena in the molecule.

The deficiencies of Becke's atomic charges can be minimized by computing $w_A(\mathbf{r})$ from topological atomic radii. The B-Top method gives atomic charges which are much more reasonable from a chemical point of view and which are fairly similar to those derived from the QTAM. Both the B-Top and QTAM charges suggest a relatively high charge transfer in the CO molecule. This behavior is general and also happens in many other molecules. On the other hand, Mod-H and MinDef schemes give atomic charges fairly close to each other and offer, in general, an image with more neutral atoms than B-Top and QTAM partitions. The Hirshfeld method provides the more neutral atoms in most cases. This fact is well-known and has been previously observed in a large variety of molecules (see ref 40 and references therein). It is noteworthy that, since the atomic functions $w_A(\mathbf{r})$ in the Becke and Hirshfeld schemes are completely independent of the details of the molecular wave functions, they produce atomic charges which are practically basis-set-independent. This feature of the Hirshfeld atomic charges was also observed in ref 40. We do not think this fact implies that they are more realistic than those of other partitions. It is simply a characteristic feature of these two

methods that, obviously, would also be desirable in all of the other cases, thus making the discussion of results easier and avoiding the inconveniences derived from the change of these results with the basis set employed in the calculation.

The total molecular dipole (μ) of a neutral molecule is given (in atomic units) by[11]

$$\mu = \sum_A Z_A \mathbf{R}_A - \int \mathbf{r}\rho(\mathbf{r})\, d\mathbf{r} \qquad (46)$$

where the electronic and nuclear position vectors $\mathbf{r}$ and $\mathbf{R}_A$ are measured from a common, arbitrary origin. Using $\mathbf{r}_A = \mathbf{r} - \mathbf{R}_A$, where $\mathbf{r}_A$ is the electronic position vector with respect to nucleus A, one can transform μ to give

$$\mu = \sum_A Q_A \mathbf{R}_A - \sum_A \int \mathbf{r}_A\, \rho_A(\mathbf{r})\, d\mathbf{r} = \mu_c + \mu_a \qquad (47)$$

where Q_A is the total (nuclear plus electronic) charge of atom A. The first term in eq 47, μ_c, is the contribution from the interatomic charge transfer, while the second, μ_a, arises from the polarization of the individual atomic distributions. Both are important in determining μ, although μ_c usually dominates when there is a significant charge transfer. In diatomic molecules, the polarity of μ_c can be generally inferred from the electronegativities of both atoms. In the CO molecule, it must be clearly of the form $C^{\delta+}O^{\delta-}$, which is the one exhibited in Table 1 by all except the Becke partition. This result does not contradict the fact that total polarity in CO is the opposite (i.e., $C^{\delta-}O^{\delta+}$, as is now the consensus from high-level calculations, and happens also in our HF calculation) since the polarization contribution (μ_a), in this case, opposes and dominates μ_c.

B. Energy Partition in H_2. The first example in which we analyze the energy partition is H_2 in the $^1\Sigma_g^+$ ground state at the CAS[2,2]/6-311G(p) level of calculation. Since the QTAM results have been discussed in detail elsewhere,[23] we will concentrate here on the comparison of these results with those obtained in the other partitions. Moreover, since H_2 is homonuclear, Becke and B-Top partitions are, in this case, equivalent. The relative values of the energy components can be rationalized in terms of the shapes of their atomic densities $\rho_A(\mathbf{r})$ and weights $w_A(\mathbf{r})$. We have depicted $\rho_A(\mathbf{r})$ for the left hydrogen of H_2 (left-H) along the internuclear axis in Figure 2. As expected, all densities are very close to each other to the left of the nucleus, differing only in the right region: the rate of decay of $\rho_A(\mathbf{r})$ is QTAM > B-Top > Mod-H > MinDef. Obviously, from the equation $\rho_A(\mathbf{r}) = \rho(\mathbf{r})\, w_A(\mathbf{r})$, the same can be said of the $w_A(\mathbf{r})$'s. The right tail of $\rho_A(\mathbf{r})$ in Figure 2 for the B-Top, Mod-H, and MinDef partitions shows a shift of QTAM electron charge from the left-H region to the right of the normal plane bisecting the H−H internuclear axis, increasing the interpenetration of the two atomic densities in the order QTAM < B-Top < Mod-H < MinDef. The charge redistribution from the neighborhood of the left-H should decrease the magnitude of V_{ne}^{AA}, directly dependent on $\rho_A(\mathbf{r})$. In the same way, the more concentrated the left-H $\rho_A(\mathbf{r})$ is on its own nucleus, the greater the value of V_C^{AA} will be. According to these arguments, the magnitudes of V_{ne}^{AA} and V_C^{AA} should decrease in the order QTAM > B-Top > Mod-H > MinDef. On the other hand, V_τ^{AA} (τ = xc, X, corr) values are given by eq 26 (halved and with $A = B$) with $\rho_2^{\tau,AA}(\mathbf{r}_1,\mathbf{r}_2) = w_A(\mathbf{r})^2\, \rho_2(\mathbf{r}_1,\mathbf{r}_2)$. From our previous discussion on the behavior of the $w_A(\mathbf{r})$'s in the different partitions, we expect that the magnitudes of these three energy components (all of them negative) also decrease in the above order.

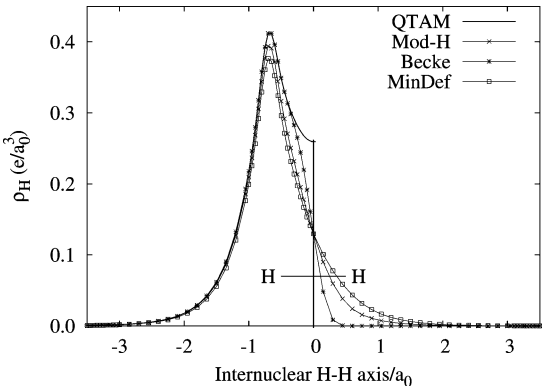


Figure 2. CAS[2,2]/6-311G(p) atomic density for the left atom of H_2 along the internuclear axis. Left and right hydrogens are at −0.7 and +0.7 a_0, respectively. Labels Becke, MinDef, Mod-H, and QTAM have been defined in the text.

The energy components at the experimental geometry (R_{exp} = 0.7414 Å) are collected in Table 2 and fully confirm the above predictions. The last row in this table collects the binding energy computed by using eq 32. In the four cases, it differs from the analytical value (eq 31) by less than 0.2 mhartree. This number may, thus, be taken as an estimate of the numerical error in the integrations.

Let us now focus on the intra-atomic properties in Table 2 (first block). As we can see, the atomic kinetic energy T^A is the same in the four partitions. Although this property suffers (except in the QTAM decomposition scheme) from the nonuniqueness of the kinetic energy density, the total molecular kinetic energy density ($\hat{T}$) is well-defined in all the schemes. Consequently, in homonuclear diatomic molecules, where both atoms are equivalent by symmetry, T^A is necessarily equal to half the total kinetic energy.

Since V_{ee}^{AA} (see eq 28) is dominated by the Coulomb part, its magnitude shows the same trend as V_C^{AA}, decreasing in the order QTAM > B-Top > Mod-H > MinDef. Nevertheless, the differences between the V_{ee}^{AA} values in the different partitions are an order or magnitude smaller than those obtained for the V_{ne}^{AA}. Since T^A is the same in all the partitions because of symmetry reasons, we conclude that V_{ne}^{AA} is the main factor determining the differences between the net energies in the different partitions.

The final intra-atomic balance gives E_{def}^A values that increase according to the sequence QTAM < B-Top < Mod-H < MinDef. This quantity is necessarily positive in homonuclear diatomic molecules,[26] so the hydrogen atom in H_2 is less destabilized with respect to the isolated state in the QTAM than in the other three partitions. In this sense, the QTAM partition is the one best preserving the atomic identity upon molecule formation. Although we have not carried out Li and Parr's division of $\rho(\mathbf{r})$, which explicitly minimizes the deformation energy,[26] it would give an E_{def}^A

Table 2. CAS[2,2]/6-311G(p) Energy Components (hartree) for H_2 at the Experimental Geometry from Different Partition Methods of $\rho(\mathbf{r})$[a]

property	QTAM	Mod-H	Becke	MinDef
T^A	0.5805	0.5805	0.5805	0.5805
V_{ne}^{AA}	−1.2278	−1.1545	−1.2149	−1.1150
V_{ee}^{AA}	0.1628	0.1590	0.1614	0.1581
V_{xc}^{AA}	−0.2367	−0.2083	−0.2299	−0.1968
V_C^{AA}	0.3995	0.3673	0.3913	0.3549
V_X^{AA}	−0.1988	−0.1818	−0.1945	−0.1753
V_{corr}^{AA}	−0.0378	−0.0265	−0.0354	−0.0216
ΔT^A	0.0807	0.0807	0.0807	0.0807
ΔV_{ne}^A	−0.2282	−0.1549	−0.2153	−0.1154
ΔV_{ee}^A	0.1628	0.1590	0.1614	0.1581
E_{net}^A	−0.4844	−0.4150	−0.4730	−0.3764
E_{def}^A	0.0154	0.0848	0.0268	0.1234
V_{ne}^{AB}	−0.5976	−0.6708	−0.6104	−0.7104
V_{ee}^{AB}	0.2994	0.3069	0.3021	0.3088
V_{cl}^{AB}	0.0423	−0.0398	0.0329	−0.0942
V_{xc}^{AB}	−0.2244	−0.2811	−0.2379	−0.3040
δ_{AB}	0.8334	0.9082	0.8502	0.9406
V_C^{AB}	0.5238	0.5880	0.5399	0.6127
V_X^{AB}	−0.2523	−0.2863	−0.2608	−0.2993
V_{corr}^{AB}	0.0279	0.0052	0.0230	−0.0046
E_{int}^{AB}	−0.1821	−0.3209	−0.2050	−0.3982
E_{bind}[b]	−0.1514	−0.1514	−0.1513	−0.1514

[a] $\Delta X^A = X^A(H_2) - X^A(H_{vac})$. Labels Becke, MinDef, Mod-H, and QTAM have been defined in the text. [b] The analytical value computed with eq 31 from the total atomic and molecular energies given by the gamess code is −0.1515 hartree.

value smaller than that of the QTAM partition. However, since our results in Table 2 nicely correlate with the general aspect of $\rho_A(\mathbf{r})$ in Figure 2, Li and Parr's $\rho_A(\mathbf{r})$ will probably be very localized on its own nucleus as the QTAM $\rho_A(\mathbf{r})$. It is also interesting to remark that, as Nalewajski et al. have recently shown,[41−43] the best transferability of the atoms from the isolated state to the molecule in the information theoretical sense is obtained when $\rho(\mathbf{r})$ is given in terms of Hirshfeld atomic densities. Preliminary results using our energy decomposition scheme fed with Hirshfeld atoms have shown, however, that, in an energetic sense, they are rather similar to the Mod-H and MinDef atoms and considerably less transferable than QTAM atoms.

We analyze now the interatomic energy components (second block in Table 2). Since each total interaction term (V_{ne}, V_C, V_{xc}, V_X, ...) is clearly independent of the partition used for $\rho(\mathbf{r})$, each of the interatomic stabilizing contributions (V_{ne}^{AB}, V_{xc}^{AB}, and V_X^{AB}) decreases now in the order MinDef < Mod-H < B-Top < QTAM, which is the opposite of what we obtained for the corresponding intra-atomic terms.

It is interesting to remark that, contrary to the intra-atomic correlation energy (V_{corr}^{AA}), which plays a stabilizing role, the interatomic correlation (V_{corr}^{AB}) destabilizes the H_2 molecule (except in the MinDef partition). The interatomic electron−electron repulsion is very similar in the four partitions, the largest difference being 5.9 kcal/mol between the QTAM and MinDef partitions. This means that, as in the intra-atomic case, the differences between the four partitions are mainly due to the electrons−nucleus interaction. In the interatomic case, V_{ne}^{AB} is, thus, the main factor causing the considerable increase (in absolute value) of the interaction energy, E_{int}^{AB}, in the order QTAM < B-Top < Mod-H < MinDef.

As shown in eq 33, E_{int}^{AB} can be decomposed in a classical (V_{cl}^{AB}) and a quantum-mechanical interaction (V_{xc}^{AB}). In most energy decomposition methods (see, for instance, ref 16), V_{cl}^{AB} is identified with the classical electrostatic interaction between the two fragments of the molecule. Since the fragment electron densities usually interpenetrate considerably, this is a highly stabilizing interaction on the order of tens or hundreds of kilocalories per mole. As we can see in Table 2, the Mod-H and MinDef partitions, based on *interpenetrating* atoms, give a negative V_{cl}^{AB} value. However, in the B-Top and QTAM partitions, V_{cl}^{AB} is a positive number. There is nothing contradictory in these results since it is trivial to show, using elementary electrostatics, that the classical interaction energy between two strictly nonoverlapping and neutral distributions of charge, one the specular image of the other, is positive. The QTAM atoms in homonuclear diatomics exactly satisfy this condition and, thus, give $V_{cl}^{AB} > 0$. However, V_{cl}^{AB} decreases as the overlap between both atomic densities increases, eventually becoming negative. The B-Top partition is one in which this overlap is not yet so strong as to give a stabilizing electrostatic interaction energy.

The above arguments lead to the following conclusions concerning the QTAM partition. Since, in homonuclear diatomics, $V_{cl}^{AB} > 0$ and $E_{def}^A > 0$, as a result of the absence of charge transfer between both atoms, the interatomic exchange-correlation term V_{xc}^{AB} is the only driving force of binding. As the interatomic overlap increases, V_{cl}^{AB} and V_{xc}^{AB} become more stabilizing, and the binding in H_2 results from a balance between E_{int}^{AB} and E_{def}^A, both quantities greater in absolute value than in the QTAM partition. The delocalization indices δ_{AB} in Table 2 increase in the order δ_{AB}(QTAM) < δ_{AB}(B-Top) < δ_{AB}(Mod-H) < δ_{AB}(MinDef). However, their values are slightly smaller than 1.0 in all the cases, which is typical of a single covalent bond. Notice that δ_{AB} is proportional to the absolute value of the corresponding interatomic exchange-correlation interaction V_{xc}^{AB}. Our results (HF and correlated) in many molecules have shown that this fact is of general validity, so the pure quantum mechanical interaction energy is a good chemical indicator of electron delocalization and of covalency. Moreover, partitioning methods of $\rho(\mathbf{r})$ based on interpenetrating atomic densities tend to give covalency indices higher than those of the QTAM partition.

Several energetic contributions to E_{bind}^A have been plotted in Figure 3 for a wide range of internuclear H−H distances. Looking at the E_{def}^A versus R_{H-H} curves, we observe that E_{def}^A is positive in all the partitions, so the interaction of both atoms has a net energy penalty, as expected. In the QTAM and B-Top partitions, E_{def}^A has a very shallow maximum near 2.5 a_0, runs through a minimum very near the equilibrium geometry, and rises very steeply when we

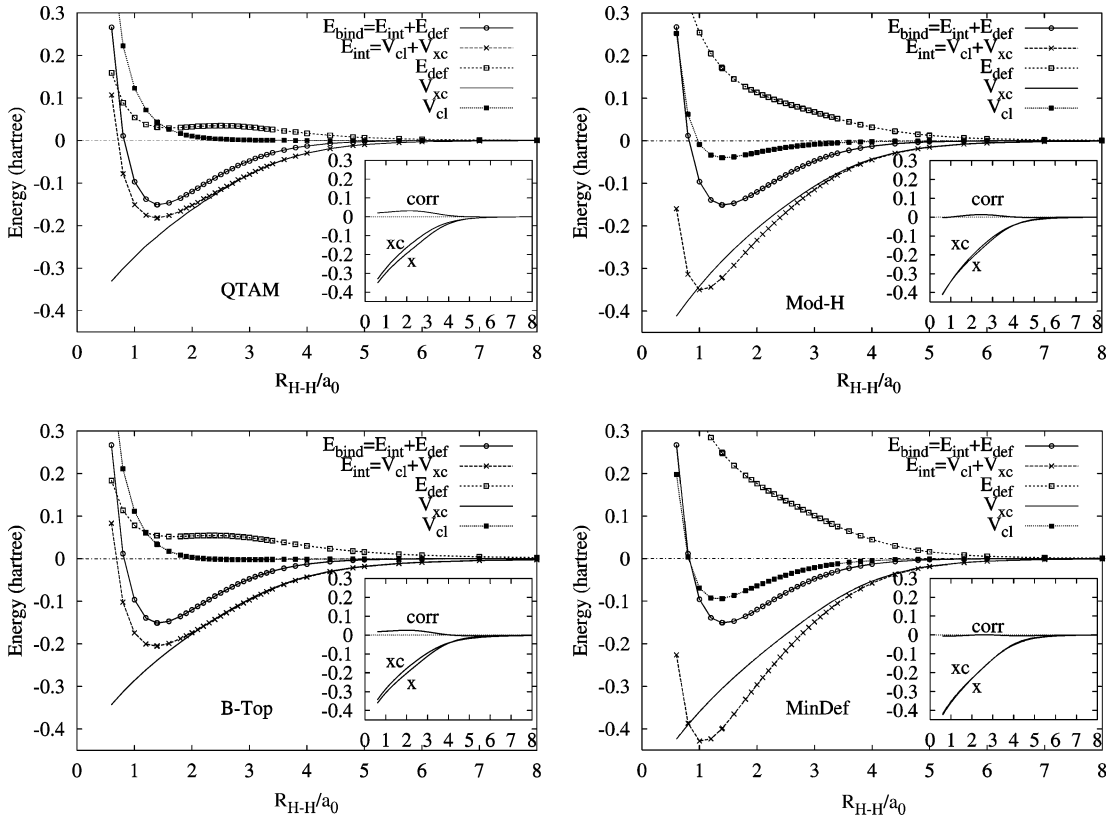


Figure 3. CAS[2,2]/6-311G(p) energy components for H_2 as a function of the internuclear distance. MinDef and Mod-H are the results using the Rico et al. and the modified Hirshfeld partitioning schemes, respectively. The insets show the splitting of V_{xc} into exchange (x) and correlation (corr) contributions.

compress the molecule in excess. However, in the Mod-H and MinDef partitions, E^A_{def} increases continuously when both atoms approach each other. Moreover, E^A_{def} at the equilibrium geometry is very small in the QTAM and B-Top partitions [9.7 (QTAM) and 16.8 (B-Top) kcal/mol]. These numbers should be compared with the corresponding E^A_{net} values [−304.0 (QTAM) and −296.8 (B-Top) kcal/mol]. There is, thus, a clear difference between the QTAM/B-Top and Mod-H/MinDef hydrogen atoms: While QTAM and B-Top hydrogen atoms in the H_2 molecule have almost the same energy that they have in vacuo, Mod-H and MinDef hydrogen atoms in the H_2 molecule are highly destabilized with respect to the isolated state. As a consequence, QTAM and B-Top binding energy curves follow faithfully those of the interaction; that is, a slight increase in deformation energy is traded for a large interaction energy.

Even more interesting is the fact that, from infinity to almost the equilibrium geometry, the interaction energy in the QTAM and B-Top partitions is practically dominated by the pure quantum mechanical interaction V^{AB}_{xc}, the classical interaction V^{AB}_{cl} being almost negligible in that regime of R_{H-H} distances. As we can see in Figure 3, the overall picture in the Mod-H and MinDef partitions is rather different. In these two cases, the binding energy curve results from the sum of two strongly attractive curves (V^{AB}_{xc} and V^{AB}_{cl}) and a strongly repulsive one (E^A_{def}). It is interesting to remark, however, that in all the cases the intercenter exchange-correlation energy, V^{AB}_{xc}, is dominated by the exchange part, which we have to understand here in the sense of Heitler−London resonance energy, and that the intercenter correlation energy, V^{AB}_{corr}, is rather small (almost negligible in the Mod-H and MinDef partitions). Besides this, V^{AB}_{corr} at the equilibrium geometry is a destabilizing contribution in all but the MinDef partition.

C. Energy Partition in N_2. We analyze in this subsection the results of our energy partition methods in the N_2 molecule. The more significant energy components for N_2 at the theoretical equilibrium geometry are gathered in Table 3 (as in H_2, the Becke and B-Top partitions are equivalent). This molecule, as any homonuclear diatomic, lacks electron charge transfer between the atoms. Each of the intra-atomic energy contributions refers, then, to the same number of electrons that a nitrogen atom has in vacuo and can, thus, be compared with its corresponding isolated value. The numbers in Table 3 give us a picture that seems to be a scaled counterpart of that of the H_2 molecule. In N_2, however, the overall numerical error in the integration is considerably larger than in H_2. Here, the binding energy computed through eq 32 is 0.4 (QTAM), 4.3 (Mod-H), 3.6 (B-Top), and 4.5 (MinDef) mhartree larger than the analytical value obtained through eq 31. In the Mod-H, B-Top, and MinDef partitions, this error is necessarily associated with the nuclear attraction and electron repulsion energies since the total kinetic energy in these schemes ($T = 2T^A = 109.1352$ hartree) reproduces with four decimal figures the exact *gamess* value ($T_{analytical} = 109.135\,238$ hartree).

As in that molecule, the different behavior of the four energy partitions in N_2 can also be understood in terms of

Table 3. CAS[10,10]/TZV(d) Energy Components (hartree) for N_2 at the Theoretical Equilibrium Geometry from Different Partition Methods of $\rho(\mathbf{r})$

property	QTAM	Mod-H	B-Top	MinDef
T^A	54.5673	54.5676	54.5676	54.5676
V_{ne}^{AA}	−129.6517	−128.9735	−129.3203	−128.7652
V_{ee}^{AA}	20.7465	20.4185	20.5689	20.3293
V_{xc}^{AA}	−6.3427	−6.2174	−6.2701	−6.1876
V_{C}^{AA}	27.0892	26.6359	26.8390	26.5169
V_{X}^{AA}	−6.0927	−6.0019	−6.0422	−5.9823
V_{corr}^{AA}	−0.2501	−0.2155	−0.2279	−0.2053
E_{net}^{A}	−54.3379	−53.9873	−54.1838	−53.8682
E_{def}^{A}	0.0609	0.4114	0.2149	0.5305
V_{ne}^{AB}	−21.9580	−22.6375	−22.2906	−22.8458
V_{ee}^{AB}	20.0432	20.7051	20.4036	20.8837
V_{cl}^{AB}	0.2216	−0.2253	0.0627	−0.4039
V_{xc}^{AB}	−0.6730	−0.9232	−0.8190	−0.9827
δ_{AB}	1.9395	2.2929	2.1583	2.3729
V_{C}^{AB}	20.7163	21.6284	21.2226	21.8664
E_{int}^{AB}	−0.4514	−1.1485	−0.7563	−1.3866
E_{bind}[a]	−0.3297	−0.3258	−0.3265	−0.3256

[a] The analytical value computed with eq 31 from the total atomic and molecular energies given by the gamess code is −0.3301 hartree.

the form exhibited by the atomic densities of both nitrogen atoms. The interpenetration of these densities increases in the order QTAM < B-Top < Mod-H < MinDef. Correspondingly, the absolute values of all the intra-atomic energy components increase in the opposite sense. The atomic deformation energy, E_{def}^{A}, in the QTAM partition is about 38 kcal/mol. This is a small number and amounts to only 0.11 of the net atomic energy. However, the Mod-H and MinDef partitions give unreasonable deformation energies (258 and 333 kcal/mol, respectively). Again, the B-Top partition, with a deformation energy equal to 135 kcal/mol, gives an intermediate result. From these numbers, it seems that one can only recognize the nitrogen atom in the QTAM partition.

As in H_2, B-Top atomic densities do not interpenetrate sufficiently as to give a negative V_{cl}^{AB} value. However, Mod-H and MinDef partitions give a strongly stabilizing classical interaction. In this sense, they are similar to other energy partition methods based on interpenetrating fragments.[16] Furthermore, the classical interaction energy, except in the B-Top partition, is not a small contribution to the total interaction energy, E_{int}^{AB}. This result contrasts with that observed in H_2, where most of the interaction was due to the exchange-correlation term, V_{xc}^{AB}.

The energy partition in N_2 has been performed in a range of N−N internuclear distances going from 1.2 to 3.6 a_0. Most of the comments concerning H_2 are also pertinent in N_2, and we have, thus, omitted the figure for brevity. From 1.8 up to 3.6 a_0, the atomic deformation energy is practically flat in the QTAM partition, whereas it continuously increases as R_{N-N} decreases in the other partitions, particularly in the Mod-H and MinDef partitions. The absolute value of the classical interaction, V_{cl}^{AB}, is always repulsive (attractive) and decreases with R_{N-N} in the QTAM (Mod-H and MinDef) partition(s). In the B-Top partition, its behavior is very similar to that found in the QTAM partition, although V_{cl}^{AB} becomes slightly attractive for R_{N-N} distances larger than 2.4 a_0. Similarly, the total interaction energy, E_{int}^{AB}, displays a minimum in the QTAM and B-Top partitions while it decreases abruptly and monotonically in the Mod-H and MinDef partitions when R_{N-N} decreases. Moreover, at R_{N-N} distances close to the equilibrium, the E_{int}^{AB} and E_{bind}^{AB} curves do not differ too much in the QTAM partition, whereas they differ greatly in the Mod-H and MinDef partitions.

In summary, we have found that, as in H_2, the QTAM partition is again the one best preserving the atomic identity in passing from the isolated atom to the molecule, followed by the B-Top, Mod-H, and MinDef partitions. Binding in the Mod-H and MinDef partitions arises, thus, as a consequence of a very delicate interplay of large (effective intra-atomic and interatomic) magnitudes, whereas in the B-Top and (more notably) QTAM partitions, it results from an interaction energy slightly contaminated by the intra-atomic destabilizing deformation energy term. Preliminary results in other homonuclear diatomics indicate that this general conclusion is valid as well.

D. Energy Partition in LiH. Let us turn to the LiH molecule. Its more relevant results at the experimental geometry are collected in Table 4. The error in the numerical integrations within the Mod-H, B-Top, and MinDef partitions is rather small (∼0.0−0.1 mhartree), while it is considerably larger in the QTAM partition (∼1.1 mhartree). The total kinetic energy (T) is 8.0119, 8.0118, 8.0118, and 8.0118 hartree in the QTAM, B-Top, Mod-H, and MinDef partitions, respectively. These numbers agree very well with the exact value (8.011874 hartree). It should be stressed, however, that T is shared between the Li and H atoms rather differently in the four schemes. For instance, T^H (QTAM) = 0.6405 hartree, while T^H takes the values 0.6249, 0.5482, and 0.7362 hartree in the B-Top, Mod-H, and MinDef partitions, respectively. Given the good agreement between T (QTAM) and the exact T value, it is clear that, in the QTAM partition, the 1.1 mhartree error in E_{bind} is due to numerical errors in the integration of the nuclear attraction and electron repulsion energies, as it happened in the N_2 molecule within the B-Top, Mod-H, and MinDef schemes.

The QTAM partition predicts that LiH is highly ionic, with atomic charges close to nominal ones. On the contrary, in the Mod-H partition, this molecule presents a relatively low ionicity, while the B-Top and MinDef partitions give intermediate results.

The rationalization of the deformation energy is not as easy as in homonuclear diatomics as a result of the electron density transfer from Li to H. Nevertheless, it is still possible to do some qualitative reasoning about its value and behavior. If the total charge of an atom would remain unchanged when it enters into a molecule, E_{def}^{A} would be strictly positive because of the variational principle. However, in heterodiatomics, one must take into account the change in E_{def}^{A} due to the CT prior to considering the term coming from the exclusive deformation of the density. In LiH, the Li atom loses a fraction (f) of an electron (different depending on

Table 4. CAS[2,2]//6-311G(p) Energy Components for LiH at the Experimental Geometry from Different Partition Methods of $\rho(\mathbf{r})$[a]

properties	QTAM	Mod-H	B-Top	MinDef
Q^{Li}	0.8912	0.4067	0.7076	0.6768
ΔT^{Li}	−0.0608	0.0314	−0.0453	−0.1566
ΔV_{ee}^{LiLi}	−0.4756	−0.1648	−0.3449	−0.3754
ΔV_{ne}^{LiLi}	0.7248	0.3036	0.5450	0.6508
V_{xc}^{LiLi}	−1.6632	−1.6865	−1.6653	−1.6314
V_{X}^{LiLi}	−1.6626	−1.6805	−1.6624	−1.6293
V_{corr}^{LiLi}	−0.0006	−0.0060	−0.0029	−0.0021
E_{def}^{Li}	0.1883	0.1702	0.1548	0.1187
Q^{H}	−0.8907	−0.4062	−0.7074	−0.6763
ΔT^{H}	0.1407	0.0484	0.1251	0.2364
ΔV_{ee}^{HH}	0.4054	0.2598	0.3532	0.3459
ΔV_{ne}^{HH}	−0.5255	−0.3241	−0.4642	−0.4224
V_{xc}^{HH}	−0.4671	−0.3232	−0.4146	−0.3808
V_{X}^{HH}	−0.4231	−0.2825	−0.3713	−0.3406
V_{corr}^{HH}	−0.0440	−0.0407	−0.0433	−0.0401
E_{def}^{H}	0.0205	−0.0159	0.0141	0.1599
V_{ne}^{LiH}	−1.8056	−1.3846	−1.6261	−1.7317
V_{ne}^{HLi}	−0.7006	−1.3846	−0.7620	−1.7317
V_{ee}^{LiH}	1.2285	1.0623	1.1493	1.1868
V_{cl}^{LiH}	−0.2394	−0.0651	−0.1499	−0.1919
$Q^{Li}Q^{H}/R_{Li-H}$	−0.2673	−0.0551	−0.1669	−0.1526
V_{xc}^{LiH}	−0.0383	−0.1591	−0.0889	−0.1567
V_{X}^{LiH}	−0.0377	−0.1605	−0.0897	−0.1536
V_{corr}^{LiH}	−0.0007	0.0014	0.0009	−0.0030
E_{int}^{LiH}	−0.2777	−0.2243	−0.2388	−0.3486
E_{bind}[b]	−0.0689	−0.0699	−0.0699	−0.0700
δ_{LiH}	0.2274	0.8002	0.5128	0.6919

[a] Atomic units are used throughout. [b] The analytical value computed with eq 31 from the total atomic and molecular energies given by the gamess code is −0.0700 hartree.

the partition) and the H atom gains that fraction of an electron. Consequently, we expect that the CT contribution to E_{def}^{Li} will be positive and on the order of $f \times$ IP, where IP is the ionization potential of Li.

Using the Q^{Li} values of Table 4 and the experimental IP of Li, we obtain for E_{def}^{Li}(CT) the values (in hartrees) 0.1747 (QTAM), 0.1387 (B-Top), 0.0797 (Mod-H), and 0.1327 (MinDef). The QTAM and B-Top numbers are reasonably close to (and smaller than) the corresponding total energy deformation values, which seems to indicate that, for Li in these two partitions, the CT effect is dominant over the effect as a result of the intrinsic charge density deformation. On the other hand, the approximation $E_{def}^{Li}(CT) = f \times IP$ is even qualitatively wrong in the MinDef partition, for this number is greater than E_{def}^{Li}. Concerning the H atom, since E_{def}^{H}(CT) has to be negative whereas the E_{def}^{H} value due to the intrinsic charge density deformation has to be positive, both quantities tend to cancel out and one should expect small total E_{def}^{H} values. The numbers in Table 4 confirm this result except in the MinDef partition, where E_{def}^{H} is too great. This is due to the kinetic energy of this atom in this partition, as shown by the ΔT^{H} value in Table 4.

Other remarkable facts relative to the intra-atomic components are the following. Most of the exchange-correlation energy is, in fact, exchange. It is interesting to remark that V_{corr}^{HH} is considerably larger (absolute value) than V_{corr}^{LiLi}. As corresponds to a positive charge for Li and a negative charge for H, $\Delta V_{ne}^{LiLi} > 0$, whereas $\Delta V_{ne}^{HH} < 0$. The contrary happens for the electron−electron repulsion; that is, $\Delta V_{ee}^{LiLi} < 0$, while $\Delta V_{ee}^{HH} > 0$.

We finally analyze the interatomic energies. We observe in Table 4 that most of the classical electrostatic energy, V_{cl}^{LiH}, can be recovered from just the point-charge term, $Q^{Li} \times Q^{H}/R_{Li-H}$. The rest of the classical interaction,[23] which collects the classical multipolar (other than charge−charge) and overlap (in the sense of ref 24) contributions, is positive in the QTAM and B-Top partitions but negative in the Mod-H and MinDef partitions. The exchange-correlation term, V_{xc}^{LiH}, correlates very well with the delocalization index, δ_{LiH}, and is very similar to the pure exchange contribution, V_{X}^{LiH}. The interatomic correlation energy, V_{corr}^{LiH} (as it also happened with the intra-atomic ones), is thus very small, in agreement with conventional wisdom. We must notice that the relative contribution of V_{cl}^{LiH} and V_{xc}^{LiH} to the total interaction energy, E_{int}^{LiH}, is very different in the four partitions. Thus, QTAM and, to a smaller degree, B-Top agree with the traditional image of ionic bonding (large and negative classical interaction with small positive contributions from overlap repulsion, here, corresponding to E_{def}). Furthermore, E_{int}^{LiH} also differs considerably in the four cases, and in the MinDef partition, this quantity is noticeably more negative than in the other three. Finally, since the binding energy, E_{bind}^{LiH}, is the same in all the cases (except for numerical errors in the integrations), it is clear that E_{int}^{LiH} and $E_{def}^{Li} + E_{def}^{H}$ contribute to E_{bind}^{LiH} in a rather different form in the four partitions.

IV. Summary and Conclusions

A molecular energy decomposition scheme based on the Li and Parr[26] partition of the nondiagonal first-order and diagonal second-order density matrices is proposed, which splits the total energy into intra-atomic and interatomic components. The method can be applied with both single-determinant (HF) or multideterminant wave functions, is independent of the details concerning the determination of the molecular wave function, and can deal equally well with different partitions of the electron density $\rho(\mathbf{r})$ into atomic contributions. Several of these partitions of $\rho(\mathbf{r})$, including the one based on the atoms provided by the quantum theory of atoms in molecules,[23] have been numerically explored by computing the energy components of H_2, N_2, and LiH molecules.

In H_2 and N_2, where electron charge transfer is absent, we have found that the relative values of the different intra-atomic and interatomic energy components are almost exclusively determined by the shape of each atomic density. Nonoverlapping atomic densities tend to give (absolute value) smaller intra-atomic and interatomic energy components than overlapping densities. The larger the overlap, the more difficult it is to recognize the original (i.e., isolated state)

atoms within molecules. In this sense, the QTAM partition, discussed in full detail in ref 23 for a representative set of molecules, is specially useful, for it provides a very appealing picture of chemical binding: atoms, relatively unchanged with respect to their in vacuo states, simply interact to form the molecule. In this partition, the atomic deformation energy is, thus, relatively small. Moreover, the large exchange-correlation interaction is the only one responsible for binding, since the total classical interaction is overall repulsive and, consequently, tends to destabilize the molecule. The image provided by the QTAM partition is certainly close to that successfully used over the years in semiempirical atomistic simulations,[44−46] in which the atomic self-energies are assumed to be approximately constant and the focus is put on the interatomic energies. Energy partitions based on strongly interpenetrating atoms tend to destroy this conventional image, as the final value of the total binding energy is the consequence of a delicate balance between intra-atomic and interatomic interactions, both of them considerably larger than the binding energy itself. We are currently applying the present energy partition method using several partitions of $\rho(\mathbf{r})$ to other homonuclear diatomics. We do not expect, however, to arrive to conclusions qualitatively different from those obtained in H_2 and N_2 molecules.

Concerning heteronuclear diatomics, the atom loosing charge has a positive and large deformation energy regardless of the partition used. This is due to both its loss of electron population and the intrinsic deformation of its atomic density with respect to the isolated state. Since both effects tend to cancel out in the negatively charged atom, it usually (but not always) has an absolute value of the deformation energy smaller than that of the positively charged atom. Contrary to homonuclear diatomics, the classical interaction energy plays a stabilizing role in the binding. Moreover, in all the partitions, most of this interaction corresponds to the point-charge interaction. The image of binding in homonuclear diatomics, almost exclusively due to the quantum-mechanical exchange-correlation interaction, is no longer valid. Here, both the classical and the pure quantum-mechanical components are relevant in understanding the binding.

Because of the existence of charge-transfer effects, a detailed comparison of the energy components obtained with different partition schemes of $\rho(\mathbf{r})$ in heteronuclear diatomics and molecules with more that two atoms is considerably more difficult than in homonuclear diatomics. To deepen our knowledge about this comparison, we are currently working on developing a sensible method to split the energy components (both intra- and interatomic) into two different contributions: a first one due to charge-transfer effects and a second one due to the intrinsic deformation of each atomic density.

Acknowledgment. Financial support from the MCyT (Project BQU2003-06553) is gratefully acknowledged.

References

(1) Fukui, K. *Acc. Chem. Res.* **1971**, *4*, 57.

(2) Fukui, K. *Theory of Orientation and Stereoselection*; Springer-Verlag: Berlin, 1975.

(3) Woodward, R. B.; Hoffmann, R. *The Conservation of Orbital Symmetry*; Verlag Chemie: Weinheim, Germany, 1970.

(4) Jeziorski, B.; Moszynski, R.; Szalewicz, K. *Chem. Rev.* **1994**, *94*, 1887.

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

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

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

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

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

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

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

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

(13) Day, P. N.; Jensen, J. H.; Gordon, M. S.; Webb, S. P.; Stevens, W. J.; Krauss, M.; Garmer, D.; Basch, H.; Cohen, D. *J. Chem. Phys.* **1996**, *105*, 1968.

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

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

(16) Bickelhaupt, F. M.; Baerends, E. J. *Rev. Comput. Chem.* **2000**, *15*, 1.

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

(18) Frenking, G.; Wichmann, K.; Fröhlich, N.; Loschen, C.; Lein, M.; Frunzke, J.; Rayón, V. M. *Coord. Chem. Rev.* **2003**, *238*, 55.

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

(20) Mayer, I. *Chem. Phys. Lett.* **2003**, *382*, 265.

(21) Mayer, I.; Salvador, P. *Chem. Phys. Lett.* **2003**, *383*, 368.

(22) Alcoba, D. R.; Torre, A.; Lain, L.; Bochicchio, R. C. *J. Chem. Phys.* **2005**, *122*, 074102.

(23) Blanco, M. A.; Martín Pendás, A.; Francisco, E. *J. Chem. Theory Comput.* **2005**, *1*, 1096.

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

(25) Martín Pendás, A.; Francisco, E.; Blanco, M. A. *J. Comput. Chem.* **2004**, *26*, 344.

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

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

(28) Baerends, E. J.; Gritsenko, O. V. *J. Phys. Chem. A* **1997**, *101*, 5383.

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

(30) Rico, J. F.; López, R.; Ramírez, G. *J. Chem. Phys.* **1999**, *110*, 4213.

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

(32) Bader, R. F. W. *Monatsh. Chem.* **2005**, *136*, 819.

(33) Kullback, K.; Leibler, R. A. *Ann. Math. Stat.* **1951**, *22*, 79.

(34) Becke, A. D. *J. Chem. Phys.* **1988**, *88*, 2547.

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

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

(37) Popelier, P. L. A.; Jouber, L.; Kosov, D. S. *J. Phys. Chem. A* **2001**, *105*, 9254.

(38) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347.

(39) Koga, T.; Watanabe, S.; Kanayama, K.; Yasuda, R.; Thakkar, A. J. *J. Chem. Phys.* **1995**, *103*, 3000.

(40) Guerra, C. F.; Handgraaf, J. W.; Baerends, E. J.; Bickelhaupt, F. M. *J. Comput. Chem.* **2003**, *25*, 189.

(41) Nalewajski, R. F.; Parr, R. G. *Proc. Natl. Acad. Sci.* **2000**, *97*, 8879.

(42) Nalewajski, R. F.; Parr, R. G. *J. Phys. Chem. A.* **2001**, *105*, 7391.

(43) Parr, R. G.; Ayers, P. W.; Nalewajski, R. F. *J. Phys. Chem. A.* **2005**, *109*, 3957.

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

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

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

CT0502209

JCTC Journal of Chemical Theory and Computation

Hydrolysis of the Anticancer Drug Cisplatin: Pitfalls in the Interpretation of Quantum Chemical Calculations

Justin Kai-Chi Lau* and Dirk V. Deubel*

ETH Zurich, USI Campus, Computational Science, Via Guiseppe Buffi 13, CH-6900 Lugano, Switzerland

Received September 13, 2005

Abstract: All three hydrolysis reactions of the anticancer drug cisplatin, *cis*-$[Pt(NH_3)_2Cl_2]$, including the acidity constants (pK_a) of the aqua complexes have been compared using a combined density functional theory (DFT) and continuum dielectric model (CDM) approach. The calculations predict very similar activation barriers (25−27 kcal/mol) and reaction free energies (0−2 kcal/mol) for each of the three hydrolysis reactions. The predicted relative free energies of both Pt(II) and Ru(II) anticancer complexes agree well with available experimental values. However, our calculated data strongly disagree with several recent computational studies that predicted the second and third hydrolysis to be thermodynamically highly unfavorable and thus would have ruled out the involvement of *cis*-$[Pt(NH_3)_2(OH_2)_2]^{2+}$ and *cis*-$[Pt(NH_3)_2(OH_2)(OH)]^{+}$ in the mode of action of the drug. This controversy can be resolved by the fact that former computational predictions of activation and reaction free energies in solution were based on second-shell reactant adducts and product adducts, which are the correct endpoints of the intrinsic reaction coordinate in vacuo but artifacts in aqueous solution.

Objective

Aiming to predict potentially active species in the mode of action of the anticancer drug cisplatin (*cis*-$[Pt(NH_3)_2Cl_2]$),[1] many quantum chemical studies have focused on the hydrolysis of one or both platinum−chloro bonds of the drug (Figure 1).[2−4] Most computational work arrived at the conclusion that both the second[2i] and third[2j] hydrolysis are strongly endothermic and thus neither *cis*-$[Pt(NH_3)_2(OH_2)_2]^{2+}$ nor *cis*-$[Pt(NH_3)_2(OH_2)(OH)]^{+}$ are involved in the anticancer activity of cisplatin. Such conclusions are traditionally[2b,c] based on the calculated energy of the transition state (TS) and a product adduct (PA) relative to the energy of a reactant adduct (RA). In RA and PA, a water molecule and chloride, respectively, are located in the second coordination shell of the metal (Figure 2). Because the intrinsic reaction coordinate[5] calculated in vacuo does end at such adducts and the activation barriers reported in most papers appear to be in good agreement with experimental values, recent studies on cisplatin hydrolysis and related reactions in aqueous solution have uncritically inherited this strategy.

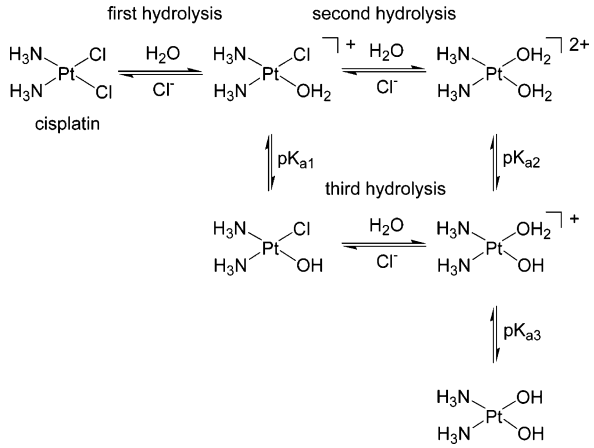


Figure 1. Cisplatin hydrolysis.

To compare for the first time all three hydrolysis reactions of cisplatin including the acidity constants (pK_a) of the aqua complexes, we have performed a combined density functional

* Corresponding author e-mail: kai.lau@phys.chem.ethz.ch (J.K.-C.L.) and metals-in-medicine@phys.chem.ethz.ch (D.V.D.).

10.1021/ct050229a CCC: $33.50
Published on Web 12/13/2005

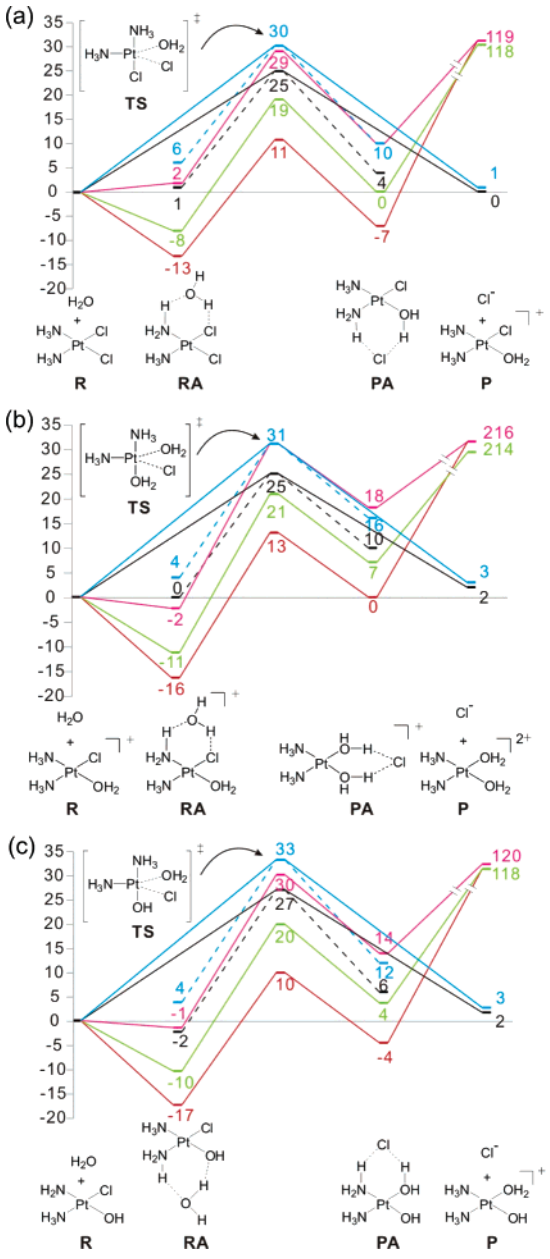


Figure 2. Calculated reaction profile (in kcal/mol) for the (a) first, (b) second, and (c) third hydrolysis of cisplatin. Red: Energies at B3LYP with small basis set in vacuo. Green: Improved energies with large basis set in vacuo. Purple: Free energies in vacuo. Blue: Free energies in solution, with Poisson−Boltzmann calculations. Black: Free energies in solution, with Poisson−Boltzmann calculations, Wertz correction included. Dashed lines: Reactant adduct (RA) and product adduct (PA) are used as a reference in aqueous solution.

theory (DFT) and continuum dielectric model (CDM) study. Our calculations suggest second-shell adducts to be artifacts from the calculations in vacuo, calling for a critical reassessment of former computational results.

Results and Discussion

Figure 2a displays the calculated reaction profile at the B3LYP level[6,7] for the *first hydrolysis* of cisplatin, including the separated reactants (R), the reactant adduct (RA), the transition state (TS), the product adduct (PA), and the separated products (P). The energy of the TS relative to the reactants (R) increases by 8 kcal/mol when instead of a common double-ζ basis set (red) a triple-ζ basis set is used (green), indicating that the values reported in some former works are altered by severe basis-set superposition errors. Entropic corrections at room temperature (purple) increase the relative energy of RA, TS, and PA by 10 kcal/mol. Consideration of solvation free energies (blue) with Poisson−Boltzmann calculations decreases the reaction free energy from 119 to 1 kcal/mol and increases the free energy of RA from 2 to 6 kcal/mol.[8] Hence, the reactant adduct (RA) would be predicted in solution to be significantly less stable than the separated reactants (R), casting doubt on the physical basis of taking RA as the reference state. Note that a Car-Parrinello study with a larger number of explicit solvent molecules did not give evidence for such adducts,[9] i.e., the attacking water molecule comes from bulk solution.

The calculated activation free energy (Figure 2a, blue) for the first hydrolysis (30 kcal/mol) relative to the separated reactants (R) is larger than experimental values (24 kcal/mol).[10,11] We believe that the continuum dielectric models do not properly consider the changes of solvation entropy in bimolecular reactions. According to Wertz and others,[12] various molecules lose a constant fraction (∼0.5) of their entropy, when they are dissolved in water. Therefore, the solvation entropy of each species including that of the TS may be assumed to be half of the entropy in vacuo with the opposite sign. With this empirical correction (Figure 2a, black),[13] the predicted activation barrier (25 kcal/mol) is in good agreement with the experimental values. Furthermore, the reactant adduct (RA) is now (Figure 2a, black) approximately as stable as the separated reactants (R). This result is very convincing, because RA and R represent the same metal complex dissolved in water. Note that the experimental activation barrier would be reproduced as well by a poor approach (red) that (i) uses inappropriately the reactant adducts (RA) as the reference, (ii) suffers from basis-set superposition errors, (iii) neglects entropic corrections, and (iv) neglects solvation effects.

Analogous calculations for the *second hydrolysis* (Figure 2b, black) arrive at relative free energies for the reactants (set to 0), TS (25 kcal/mol), and products (2 kcal/mol) that are remarkably similar to those of the first hydrolysis step. In contrast, the second hydrolysis reaction would have been predicted to be 12 kcal/mol endothermic, if the reactant adduct (RA) and product adduct (PA) had been taken into account (Figure 2b, blue, dashed lines). This result would have suggested *cis*-$[Pt(NH_3)_2(OH_2)_2]^{2+}$ not to be involved at all in the mode of action of cisplatin. Such an interpretation would have ignored the result that the product adduct (PA) is 8 kcal/mol less stable than the separated products (P), i.e., a fully solvated chloride ion is significantly more stable than a chloride in the second coordination shell of the aqua complex.

As an alternative to the second hydrolysis, *cis*-$[Pt(NH_3)_2(OH_2)Cl]^{+}$ may be deprotonated first, and then the Pt−Cl bond of *cis*-$[Pt(NH_3)_2(OH)Cl]$ may be hydrolyzed, herein denoted *third hydrolysis* (Figure 2c, black). For the third

Table 1. Comparison of Calculated and Experimental Activation Free Energies (ΔG_a; in kcal/mol) and Reaction Free Energies (ΔG_r; in kcal/mol) for the Hydrolysis of Pt−Cl Bonds Anticancer Complexes and Absolute pK_a Values of the Aqua Complexes[a]

metal complex		calc	exp	exp ref
cis-$[Pt(NH_3)_2Cl_2]$	ΔG_a	24.9	23.8; 24.1	c; d
cis-$[Pt(NH_3)_2Cl_2]$	ΔG_r	0.1	4.2; 3.6	c; d
cis-$[Pt(NH_3)_2(OH_2)Cl]^+$	ΔG_a	25.3	23.3	d
$[Ru(Ar)(en)Cl]^+$ [b]	ΔG_a	20.7	21.4	b
$[Ru(Ar)(en)Cl]^+$ [b]	ΔG_r	1.1	3.2	b
cis-$[Pt(NH_3)_2(OH_2)Cl]^+$ (pK_{a1})	pK_a	7.8	6.41	e
cis-$[Pt(NH_3)_2(OH_2)_2]^{2+}$ (pK_{a2})	pK_a	8.3	5.37	e
cis-$[Pt(NH_3)_2(OH_2)(OH)]^+$ (pK_{a3})	pK_a	9.5	7.21	e
$[Ru(Ar)(en)(OH_2)]^{2+}$ [b]	pK_a	9.8	7.71	b

[a] A difference of 1 pK_a unit reflects a free energy difference of $RT\ln 10 = 1.36$ kcal/mol. [b] en = 1,2-diaminoethane. Ar = η^6-benzene (calc), η^6-biphenyl (exp). Reference 14b. [c] Coe, J. S. *MTP Int. Rev. Sci.: Inorg. Chem., Ser. 2* **1974**, 45. [d] Perumareddi, J. R.; Adamson, A. W. *J. Phys. Chem.* **1978**, *72*, 414. [e] Reference 14a.

hydrolysis, we predict an activation free energy (27 kcal/mol) that is slightly higher than the barriers for the first two hydrolysis steps, indicating that *cis*-$[Pt(NH_3)_2(OH_2)(OH)]^+$ may form more likely via deprotonation of the second hydrolysis product. The theoretical prediction of the pK_a values of the three aqua complexes of cisplatin[14a] presented in Table 1 corroborates the remarkable absolute accuracy of ~4 kcal/mol of the quantum chemical approach, while the relative accuracy appears to be even better.

The former unisonous prediction of a strongly endergonic second and third hydrolysis—the most recent papers suggested reaction free energies of 12[2i] and 8 kcal/mol,[2j] respectively—would strongly contradict the experimental detection of diaqua and hydroxo species more than two decades ago.[15] Today it is still controversial whether *cis*-$[Pt(NH_3)_2(OH_2)_2]^{2+}$ and *cis*-$[Pt(NH_3)_2(OH_2)(OH)]^+$ are responsible for the anticancer activity, in addition to *cis*-$[Pt(NH_3)_2(OH_2)Cl]^+$.[16] For instance, the rate constants for the reaction of cisplatin derivatives with GG and AG moieties of double-stranded oligonucleotides suggest the diaqua species to be the actually active species.[16] The current work is the first theoretical study on cisplatin hydrolysis that supports this possibility, together with recent theoretical studies on the reactivity of cisplatin hydrolysis products with the nucleobases.[17] The question as to whether reactant adducts play a role in DNA binding remains controversial.[17−19] In this context, it is interesting to note the experimental detection of weak noncovalent interactions of *cis*-$[Pt(NH_3)_2(OH_2)_2]^{2+}$ and oligonucleotides prior to the reaction,[20] but their structure in aqueous solution and their impact on the rates of binding to DNA in this medium has not yet been clarified.

Computational Details

The geometries of molecules and transition states (TS) were optimized at the gradient-corrected DFT level using the 3-parameter fit of exchange and correlation functionals of Becke (B3LYP),[6] which includes the correlation functional of Lee, Yang, and Parr (LYP),[7] as implemented in Gaussian 98.[21] The LANL2DZ ECP's[22] and valence-basis sets were used at platinum, and the 6-31G(d,p) basis sets were used at the other atoms.[23] This basis-set combination is denoted II. Vibrational frequencies were also calculated at B3LYP/II. The structures reported are either minima (NIMAG = 0) or transition states (NIMAG = 1) on the potential energy surfaces. Improved total energies were calculated at the B3LYP level using the same ECP and valence-basis set at the metal, but totally uncontracted and augmented with Frenking's set of f functions,[24] together with the 6-311+G(3d) basis sets at chlorine and the 6-311+G(d,p) basis sets at the other atoms. This basis-set combination is denoted III+. Activation and reaction free energies (ΔG_a, ΔG_r) were calculated by adding corrections from unscaled zero-point energy (ZPE), thermal energy, work, and entropy evaluated at the B3LYP/II level at 298.15 K, 1 atm to the activation and reaction energies (ΔE_a, ΔE_r), which were calculated at the B3LYP/III+//II level. We found a good agreement between B3LYP and CCSD(T) relative energies (see the Supporting Information), which is not unexpected.[2j,25] Additional calculations were performed using the Stuttgart-Dresden-Bonn ECP[26] and improved basis sets,[27] which gave relative energies very similar to those obtained using LANL2DZ.

Solvation free energies G_{solv}^{ϵ} of the structures optimized at the B3LYP/II level were calculated by Poisson−Boltzmann (PB)[8] calculations with a dielectric constant ϵ of the dielectric continuum that represents the solvent. The PB calculations were performed at the B3LYP level using the LACVP** basis set on platinum, the 6-31+G* basis on oxygen, and the 6-31G** basis set on the other atoms as implemented in the Jaguar 5 program package.[28] The continuum boundary in the PB calculations was defined by a solvent-accessible molecular surface with a set of atomic radii for H (1.150 Å), C (1.900 Å), N (1.600 Å), O (1.400 Å), S (1.900 Å), Cl (1.974 Å), and Pt (1.377 Å).[29] pK_a predictions were carried out using a thermodynamic cycle,[30] $\Delta G^{\epsilon} = \Delta G^1 + G_{solv}^{\epsilon}(H^+) + G_{solv}^{\epsilon}(A^-) - G_{solv}^{\epsilon}(A)$ and $pK_a^{\epsilon} = \Delta G^{\epsilon}/RT\ln 10$, where ΔG^1 and ΔG^{ϵ} are the reaction free energies of the reaction, $AH \rightarrow A^- + H^+$, in vacuo and at a dielectric constant $\epsilon = 80.37$ for water, respectively, $G_{solv}^{\epsilon}(X)$ is the solvation free energy of species AH or A^- at ϵ obtained via PB calculations, R is the ideal gas constant, and T is the temperature (298.15 K). Experimental values have been used for the hydration free energy $G_{solv}^{\epsilon}(X)$ of small molecules and ions.[31] We believe that continuum dielectric models do not consider properly the changes of solvation entropy in bimolecular reactions; comparisons with experimental values indicate that reactions of platinum complexes and palladium complexes (unpublished) are systematically about ~6 kcal/mol too high. According to Wertz and others,[12] various molecules lose a constant fraction (approximately 0.5) of their entropy, when they are dissolved in water. All free energies in solution except that of the H^+ ion in solution were modified by an entropic term that is half (0.5) of the entropy in vacuo, with the opposite sign. This empirical correction has led to predicted pK_a values of platinum aqua complexes as well as reaction and activation free energies for the hydrolysis of metal complexes that are in good agreement with experimental values (Table 1).

Acknowledgment. Dedicated to Prof. Michele Parrinello on the occasion of his 60th birthday. This work has been supported by the Swiss National Science Foundation, the Fonds der Chemischen Industrie, Germany, the Bundesministerium for Bildung and Forschung, Germany, and the Swiss National Computing Center.

Supporting Information Available: Calculated relative energies and free energies in vacuo and aqueous solution for the hydrolysis of the Pt−Cl bonds of cisplatin, *cis*-[Pt$(NH_3)_2Cl_2$], *cis*-[Pt$(NH_3)_2(OH_2)Cl]^+$, and *cis*-[Pt$(NH_3)_2(OH)$-Cl] (Table S-1). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) (a) *Cisplatin*; Lippert, B., Ed.; Wiley-VCH: Weinheim, 1999. (b) Guo, Z. J.; Sadler, P. J. *Angew. Chem., Int. Ed.* **1999**, *38*, 1512. (c) Fuertes, M. A.; Alonso, C.; Pérez, J. M. *Chem. Rev.* **2003**, *103*, 645. (d) Reedijk, J. *P. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 3611. (e) Jakupec, M. A.; Galanski, M.; Keppler, B. K. *Rev. Physiol. Bioch. Pharmacol.* **2003**, *146*, 1. (f) *Metal Ions in Biological Systems*; Sigel, A., Sigel, H., Eds.; Marcel Dekker: New York, 2004; Vol. 42. (g) Wang, D.; Lippard, S. J. *Nature Rev. Drug Discuss.* **2005**, *4*, 307.

(2) (a) Burda, J. V.; Zeizinger, M.; Sponer, J.; Leszczynski, J. *J. Chem. Phys.* **2000**, *113*, 2224. (b) Chval, Z.; Sip, M. *J. Mol. Struct. (THEOCHEM)* **2000**, *532*, 59. (c) Zhang, Y.; Guo, Z. J.; You, X. Z. *J. Am. Chem. Soc.* **2001**, *123*, 9378. (d) Zeizinger, M.; Burda, J. V.; Sponer, J.; Kapsa, V.; Leszczynski, J. *J. Phys. Chem. A* **2001**, *105*, 8086. (e) Cooper, J.; Ziegler, T. *Inorg. Chem.* **2002**, *41*, 6614. (f) Raber, J.; Llano, J.; Eriksson, L. A. In *Quantum Medicinal Chemistry*; Carloni, P., Alber, F., Eds.; Wiley-VCH: Weinheim, 2003; p 113. (g) Robertazzi, A.; Platts, J. A. *J. Comput. Chem.* **2004**, *25*, 1060. (h) Burda, J. V.; Zeizinger, M.; Leszczynski, J. *J. Chem. Phys.* **2004**, *120*, 1253. (i) Raber, J.; Zhu, C.; Eriksson, L. A. *Mol. Phys.* **2004**, *102*, 2537. (j) Burda, J. V.; Zeizinger, M.; Leszczynski, J. *Comput. Chem.* **2005**, *26*, 907.

(3) An "inverted" hydration of some Pt(II) complexes with Pt- - - H−O−H bonds was predicted as well: (a) Kozelka, J.; Berges, J.; Attias, R.; Fraitag, J. *Angew. Chem., Int. Ed.* **2000**, *39*, 198. (b) Berges, J.; Caillet, J.; Langlet, J.; Kozelka, J. *Chem. Phys. Lett.* **2001**, *344*, 573.

(4) Deubel, D. V. *Chem. Rev.*, in preparation.

(5) Fukui, K. *Acc. Chem. Res.* **1981**, *14*, 363.

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

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

(8) Marten, B.; Kim, K.; Cortis, C.; Friesner, R. A.; Murphy, R. B.; Ringnalda, M. N.; Sitkoff, D.; Honig, B. *J. Phys. Chem.* **1996**, *100*, 11775.

(9) Carloni, P.; Sprik, M.; Andreoni, W. *J. Phys. Chem. B* **2000**, *104*, 823.

(10) Surprisingly, other widely used continuum dielectric models such as COSMO or PCM (ref 11) together with a solute cavity that was used in former studies (ref 2j) would predict an even higher activation free energy (∼35 kcal/mol) than the present Poisson−Boltzmann calculations (ref 8). For details, see the Supporting Information.

(11) Cramer, C. J.; Truhlar, D. G. *Chem. Rev.* **1999**, *99*, 2161.

(12) (a) Wertz, D. H. *J. Am. Chem. Soc.* **1980**, *102*, 5316. (b) Abraham, M. H. *J. Am. Chem. Soc.* **1981**, *103*, 6742.

(13) The Wertz correction has improved the predicted activation barriers of several bimolecular reactions; cf. ref 2e.

(14) (a) Berners-Price, S. J.; Frenkiel, T. A.; Frey, U.; Ranford J. D.; Sadler, P. J. *Chem. Commun.* **1992**, 789. (b) Wang, F.; Chen, H.; Parsons, S.; Oswald, I. D. H.; Davidson, J. E.; Sadler, P. J. *Chem. Eur. J.* **2003**, *9*, 5810.

(15) Lippard, S. J. *Science* **1982**, *218*, 1075.

(16) Selected references: (a) Bancroft, D. P.; Lepre, C. A.; Lippard, S. J. *J. Am. Chem. Soc.* **1990**, *112*, 6860. (b) Legendre, F.; Bas, V.; Kozelka, J.; Chottard, J. C. *Chem. Eur. J.* **2000**, *6*, 2002. (c) Davies, M. S.; Berners-Price, S. J.; Hambley, T. W. *Inorg. Chem.* **2000**, *39*, 5603. (d) Vinje, J.; Sletten, E.; Kozelka, J. *Chem. Eur. J.* **2005**, *11*, 3863.

(17) For example, see: Baik, M.-H.; Friesner, R. A.; Lippard, S. J. *J. Am. Chem. Soc.* **2003**, *125*, 14082.

(18) Chval, Z.; Sip, M. *Collect. Czech. Chem. Commun.* **2003**, *63*, 1105.

(19) Raber, J.; Zhu, C.; Eriksson, L. A. *J. Phys. Chem. B* **2005**, *109*, 11006.

(20) Wang, Y.; Farrell, N.; Burgess, J. D. *J. Am. Chem. Soc.* **2001**, *123*, 5576.

(21) Frisch, M. J. et al. Gaussian 98; Gaussian Inc.: Pittsburgh, PA, 1998.

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

(23) Binkley, J. S.; Pople, J. A.; Hehre, W. J. *J. Am. Chem. Soc.* **1980**, *102*, 939. (b) Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257.

(24) Ehlers, A. W.; Böhme, M.; Dapprich, S.; Gobbi, A.; Höllwarth, A.; Jonas, V.; Köhler, K. F.; Stegmann, R.; Veldkamp A.; Frenking, G. *Chem. Phys. Lett.* **1993**, *208*, 111.

(25) Sponer, J. E.; Miguel, P. J. S.; Rodriguez-Santiago, L.; Erxleben, A.; Krumm, M.; Sodupe, M.; Sponer, J.; Lippert, B. *Angew. Chem., Int. Ed.* **2004**, *43*, 5396.

(26) Andrae, D.; Haeussermann, U.; Dolg, M.; Stoll, H.; Preuss, H. *Theor. Chim. Acta* **1990**, *77*, 123.

(27) Martin, J. M. L.; Sundermann, A. *J. Chem. Phys.* **2001**, *114*, 3408.

(28) Jaguar 5.0; Schrodinger, Inc.: Portland, OR, 2000. See: Vacek, G.; Perry, J. K.; Langlois, J.-M. *Chem. Phys. Lett.* **1999**, *310*, 189. www.schrodinger.com

(29) See: (a) Rashin, A. A.; Honig, B. *J. Phys. Chem.* **1985**, *89*, 5588. (b) Gilbert, T. M.; Hristov, I.; Ziegler, T. *Organometallics* **2001**, *20*, 1183. (c) Baik, M. H.; Friesner, R. A. *J. Phys. Chem. A* **2002,** *106*, 7407. (d) Baik, M. H.; Friesner, R. A.; Lippard, S. J. *J. Am. Chem. Soc.* **2002**, *124*, 4495.

(30) Jorgensen, W. L.; Briggs, J. M.; Gao, J. *J. Am. Chem. Soc.* **1987**, *109*, 6857.

(31) (a) Barone, V.; Cossi, M. *J. Chem. Phys.* **1997**, *107*, 3210. H_2O: −6.3 kcal/mol, Cl^-: −77.0 kcal/mol. (b) Chambers, C. C.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1996**, *100*, 16385. H^+: −260.9 kcal/mol.

CT050229A

JCTC Journal of Chemical Theory and Computation

Exploring the Mechanisms of Reactions in Solution from Transition Path Sampling Molecular Dynamics Simulations

Dirk Zahn*

Max-Planck Institut für Chemische Physik fester Stoffe, Nöthnitzer Strasse 40, 01187 Dresden, Germany

Received July 19, 2005

Abstract: Recent advances in molecular dynamics simulations of rare reaction events and aggregation processes are reviewed. Therein the central focus is dedicated to employing the transition path sampling method to study reactions in solution. We describe systematic approaches for generating initial transition pathways and efficient strategies for computationally feasible exploration of further transition routes. The unprejudiced study of reaction mechanisms is illustrated for reactions in aqueous solution and other complex systems. Transition path sampling allows very detailed investigation of solvent effects. Apart from stabilization of reactant, transition, or product state ensembles, this also includes the role of the solvent as a heat bath and as a putative reaction partner. The latter issue is of particular importance for reactions in aqueous solutions, which involve proton-transfer steps that may be assisted by water molecules via the Grotthuss mechanism.

1. Introduction

Many processes in solution chemistry pose two fundamental problems to the computational chemist: the need to study complex simulation models and to overcome large energy barriers, which separate reactants from product states. Apart from these limitations, molecular dynamics simulations in principle appear perfectly suited for the investigation of reaction mechanisms at the atomistic level of detail.

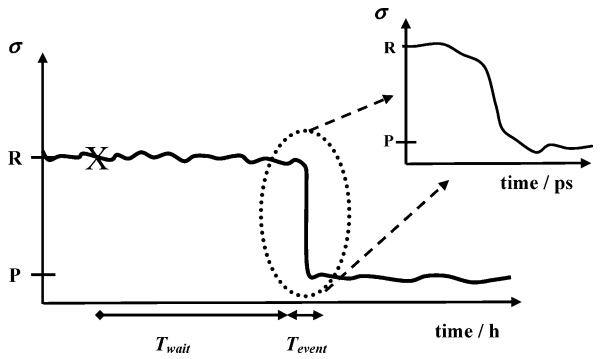
In large model systems, the computational demand not only is caused by the evaluation of a specific atomic arrangement but also is related to the immense configurational manifold arising from the large number of atoms. This particularly applies to processes, which involve the crossing of rare intermediate states. Their investigation is complicated by the need to scan a large number of possible arrangements in order to find the transition state(s). In molecular dynamics simulations this implies long 'waiting' times, before the event of interest actually occurs. These waiting periods may easily exceed the scope even of sophisticated hardware by several orders of magnitude, hence rendering the observation of many processes from direct simulation practically impossible.

In an attempt to circumvent this problem, two major approaches have emerged over the past decades. The most straightforward ansatz is to enhance the kinetics of rare events by applying elevated temperature, pressure, or strong superconcentration of a particular molecular species. While in principle this strategy helps crossing any reaction barrier, the stronger the artificial process acceleration is chosen the more careful the results have to be considered. Excessive driving may easily lead to the skipping of important intermediates or even cause the system to follow completely different mechanistic routes. Similar limitations are related to the widely used approach of applying external driving forces. This method is based on the choice of a presumed reaction coordinate. The desired process is then induced by artificial potentials or constraints, which are functions of this coordinate. As a consequence, the mechanistic analysis may only be given in terms of predefined models of the reaction coordinate. In principle this limitation may be overcome by performing several independent investigations based on various mechanistic models. However in complex systems

* Corresponding author phone: +49 (0) 351 4646 4205; fax: +49 (0) 351 4646 4002; e-mail: zahn@cpfs.mpg.de.

10.1021/ct0501755 CCC: $33.50
Published on Web 11/09/2005

Scheme 1. Illustration of an Order Parameter Plot as a Function of Time[a]



[a] While the system remains for relatively long times in the metastable reactant (R) and product (P) state regimes, the transition occurs on a much faster time scale.

the number of putative mechanistic routes typically is too large to account for all possibilities.

Recently, Chandler et al. introduced the transition path sampling (TPS) method for the molecular dynamics simulation of rare events.[1,2] This approach concentrates on a relatively short time interval in which the process of interest takes place and completely ignores the waiting period required for its observation from unconstrained simulation. As a consequence, TPS allows the study of rare events without artificial driving of the process. Moreover, no prejudicing of the reaction coordinate is needed, and the reaction mechanisms may instead be obtained as a result from the simulations. This makes TPS a very powerful tool for unbiased mechanistic investigations.

In the past few years, the TPS approach was successfully applied to a broad spectrum of processes, ranging from reactions[3−17] to conformational rearrangements[18−20] and phase transitions.[21−29] In each of these fields TPS allowed to expand the scope of molecular dynamics simulations. The present microreview describes how this method can be used for mechanistic studies of reactions in solution chemistry and identification of the role of the solvent molecules. This work was inspired by a series of recent studies, which revealed new mechanistic insights into reactions and aggregation processes in solution and demonstrated the ability of TPS to provide a very detailed picture of the solvent effect.[3−16]

2. Theory

2.1. Rare Events. Many bond breaking and formation processes are related to the crossing of large energy barriers, which separate the meta stable reactant and product states. For reactions in solution this may apply to both the reactants and the solvent molecules. The latter issue may be illustrated at the example of the association of a pair of Na^+ and Cl^- ions. In the gas phase the reaction $Na^+ + Cl^-$ (separate ions) $\rightarrow$ $Na^+ \cdot\cdot Cl^-$ (contact ion pair) is governed by the Coulomb attraction and does not exhibit an energy barrier. However, in aqueous solution the formation of a $Na^+ \cdot\cdot Cl^-$ contact ion pair requires the penetration of solvent spheres. This process implies breaking and rearrangement of hydrogen bonds, which is related to an activation energy of about 14 kJ/mol.[3]

At room temperature this barrier is only around 3 k_BT, and one has reasonable chances to observe the reaction to occur spontaneously within the picosecond to nanosecond time scale accessible to molecular dynamics simulations. However, for larger activation energies the 'waiting times' needed before the reactive events happen are considerably larger. A typical scenario of this kind is illustrated in Scheme 1. The plot shows a general order parameter σ, which reflects a quantitative measure of the reaction progress as a function of time. (For the association of $Na^+ \cdot\cdot Cl^-$ σ may be simply defined as the interionic distance). When starting a molecular dynamics simulation from an arbitrarily chosen configuration of the reaction state (indicated by the X in Scheme 1), the total time needed to observe the formation of the reaction products is given as the sum of the waiting period before the reactive event occurs and the duration of the reactive event itself.

$$T_{\text{simulation}} = T_{\text{wait}} + T_{\text{event}}$$

The larger the reaction barrier, the longer are the observed waiting times. Like this the simulation time needed for the reaction to occur may exceed the duration of the reactive event by several orders of magnitude. The key idea of the TPS approach is to only focus on the relatively short time sketch T_{event} and to largely ignore the waiting time. This concept is particularly useful for processes, which involve large energy barriers and imply long waiting times. Though T_{wait} may be very large, many of such reactions occur on a femtosecond to picosecond scale, i.e., T_{event} is sufficiently small to be covered by a molecular dynamics simulation.

While the crossing of high energy barriers accounts for a large number of rare events, slow processes may also originate to a diffusive character of the system under consideration. In an entirely diffusion controlled reaction T_{event} is large, while T_{wait} is zero. TPS then becomes quite inefficient, and other methods such as steered molecular dynamics or free energy sampling approaches are more suitable.

2.2. Transition Path Sampling in Complex Systems. While a detailed description of the TPS approach is given in refs 1, 2, 28, and 29, in this subsection we summarize the method only briefly and instead focus on more technical tricks of the trade for applying TPS to complex systems. The latter are collected from a series of studies dedicated to reactions in solution and phase transitions.[11−13,17,22−24,26]

The TPS approach represents an iterative simulation scheme, for which at least one dynamical pathway of the rare event is needed as a prerequisite. Systematic ways to generate such initial trajectories are discussed in section 2.4. Provided a first trajectory of the rare event is given, we need to define a quantitative descriptor of the reaction progress σ, as illustrated in the previous section 2.1. While the optimal reaction descriptor is of course the reaction coordinate, the latter is a priori not known. However, it is sufficient to use just one component of the reaction coordinate for describing the reaction progress. In contrast to the complete reaction coordinate, one of its components is typically very easy to find. In many cases σ is simply chosen as the distance of two atoms which undergo a bond formation or breaking in the course of the reaction.

Starting from an initial reaction pathway further trajectories are generated in an iterative procedure. For this a snapshot is taken from the preceding transition pathway, and slight changes are incorporated. This configuration variation (shooting) should be considered as a Monte Carlo step and therefore must be implemented in a manner that the simulation ensemble is conserved. This may be illustrated at the example of the microcanonical ensemble, which implies constant total energy. An easy way to realize such shooting moves is to keep the atomic positions constant and apply only momentum changes. Like this the potential energy is constant, and the conservation of the kinetic energy may be achieved from velocity rescaling. It should be noted that the momentum changes must also conserve the total momentum and angular momentum of the simulation system.

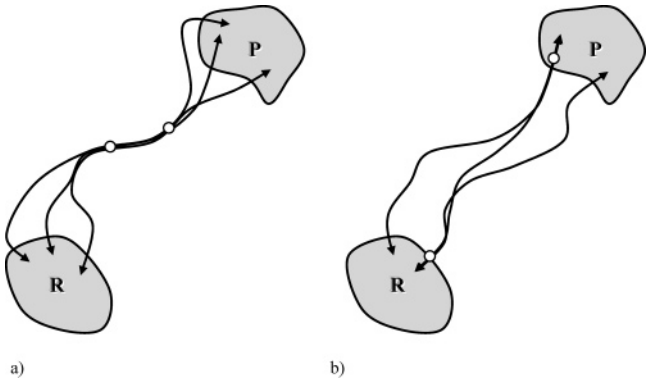
The modified configuration is then propagated in both directions of time, and the resulting trajectory is checked for the process of interest. For this purpose the reaction descriptor σ is used as a quantitative measure for the identification of pathways which go from the reactant state regime to a product state or vice versa. In case the desired event takes place, the new trajectory is chosen for generating further ones. Harvesting in an iterative manner leads to a manifold of dynamical pathways, each reflecting a possible transition route. The sampling of reaction pathways is not biased from prejudicing the reaction coordinate but instead relies on a reasonable choice of the reaction descriptor. Since σ is typically much more safe to guess than all of the components of the reaction coordinate, TPS may be considered as an unbiased method for studying reaction mechanisms.

2.3. Ergodicity and Efficient Ways of Sampling Transition Pathways. One of the most powerful features of TPS is related to its Monte Carlo type of sampling reactive trajectories. Once ergodicity is reached, the relevance of a transition route may be directly concluded from the occurrence of corresponding trajectories. For this reason the first transition pathway does not need to be a favorable one. In the course of TPS iterations the starting pathway will converge towards the preferred regions in trajectory space of reactive events.

As in all Monte Carlo simulations ergodicity of TPS, i.e., the knowledge of all reaction routes is often hard to reach. However, in solution chemistry the primary interest is related to the identification of the most preferred mechanism, while unlikely reaction pathways play a much less important role. Rather than full transition trajectory ergodicity it is therefore usually sufficient to ensure that TPS has visited the region of trajectory space corresponding to the most favored reaction mechanism.

When starting TPS from an initial trajectory which reflects an unfavorable mechanistic route, it is therefore necessary to continue the sampling iterations until the Monte Carlo moves have evolved to the most preferred class of pathways through the transition state ensemble. In a series of recent studies we elaborated some tricks of the trade how this process can be speeded up considerably.[17,24,26] To demonstrate the underlying principles, it is educative to compare two different ways of sampling transition trajectories as

Scheme 2. Different Types of Sampling Trajectories Connecting the Reactant (R) and Product (P) State Regimes[a]



[a] a) Shooting moves (o) at close distance imply poor trajectory decorrelation, though the sampling of R and P appears good. b) Trajectory modifications are applied in R and P only, resulting in good sampling of all patches of the reaction pathways.

shown in Scheme 2a,b. The illustration 2a reflects a sampling run in which the shooting moves were chosen within a short time interval compared to the total length of the transition trajectories. The shooting moves typically represent only small configuration changes. Only in the course of sufficiently long time propagation such small modifications may result in large trajectory deviations. This may be seen from the quite broad sampling of the reactant and product state regime in illustration 2a. The problematic issue indicated in Scheme 2a is related to the sampling of intermediate configurations close to the small time window in which the shooting is applied. In this region trajectory decorrelation is rather poor, and the sampling is usually far from ergodicity. This phenomenon becomes particularly inconvenient if all shooting moves are incorporated close to the transition state surface. In this case the trajectory evolution toward the most favored region of the transition state regime, i.e., the convergence of pathways to the preferred reaction mechanism requires a very large number of sampling iterations.

To avoid this limitation, subsequent shooting moves should be chosen as far as possible from each other. A very efficient approach of this kind is described in ref 26. In this work, the shooting moves are only applied at the ends of the reactive trajectories, i.e. in alternating order in the reactant and in the product state regime. Sampling in this manner proved to be very successful for fast trajectory convergence to the favored mechanistic route. This feature is illustrated in Scheme 2b. Note that the sampling of the reactant and product states is quite good, even if the shooting is applied close to these regions in trajectory space. Indeed, the shooting moves from the reactant states are used for broad sampling of the product state regime and vice versa. Like this the ends of the reaction pathways are changed in alternating order, while the intermediate sketches are rectified in each of the TPS iterations.

The use of only two states for shooting furthermore facilitates the implementation of an automated adjustment of the shooting parameters, which govern the extent of the configurational changes in each Monte Carlo move. This issue may become of considerable importance if reactant and

product state are separated by a rough free energy landscape including several barriers and local minima. In such cases the conventional TPS approach typically yields very low acceptance probabilities in both stable states. Applying only small changes during the shooting moves helps increasing the acceptance ratio, however, at the price of only small trajectory variation. To find a compromise between low acceptance ratios and large trajectory modification, we implemented an automatic procedure for adapting the shooting moves on-the-fly.[26] Therein the shooting parameters are multiplied by a factor larger than 1 in case of a successful reactive event and divided by the same number in the opposite case. After convergence of this procedure this leads to an average acceptance rate of 50%, which we suggest as a suitable compromise for computationally efficient exploration of the trajectory space of reactive events.[26]

While this two-state shooting approach is very suitable for the investigation of reaction mechanisms, the computation of rate constants requires a different sampling strategy. The original TPS scheme as developed by Chandler and co-workers[1,2] provides knowledge of the acceptance ratio as a function of an order parameter describing the reaction progress. From this one can calculate the reactive flux and the net rate constant. A particularly elegant procedure of this kind is represented by the transition interface sampling variation of TPS, which was recently introduced by Bolhuis and co-workers.[30,31]

Regardless of how the shooting moves are implemented, the checking of trajectory decorrelation and pathway convergence to the favored reaction route(s) is of vital importance for a proper mechanistic analysis. For this purpose, the Lyaponov coefficient represents a quantitative measure for the investigation of trajectory decorrelation.[28] Another approach is to start TPS from an unfavorable reaction route and count the number of sampling iterations needed for trajectory evolution to the most favored reaction mechanism. A particularly robust convergence check may be achieved by starting several independent sets of TPS simulations, each starting from different initial pathways which correspond to different mechanistic routes.[22,24] Evolution of all sets of TPS iterations to the same class of trajectories offers a quite evident proof of convergence.

Apart from running straightforward TPS iterations, one may also take use of special sampling techniques established for enhancing ergodicity in standard Monte Carlo simulations. Examples for such approaches are parallel tempering,[28,29] Wang-Landau sampling,[32] and biased TPS.[17]

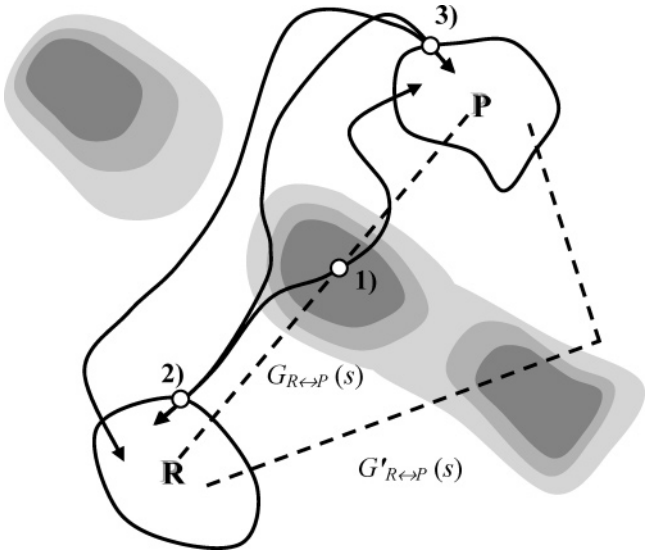
2.4. Preparation of the Initial Trajectory of a Reactive Event. In some cases the initial trajectory needed as a prerequisite for TPS iterations can be prepared by modeling a putative intermediate from intuition and propagation in both directions of time. However, for complex processes more systematic approaches may be much more efficient. We developed such a strategy, which appears quite flexible and was successfully applied in a large variety of simulation studies.[13,22−27] As a starting point, a geometric model $G_{R\leftrightarrow P}$ in real space connecting an arbitrarily chosen reactant to a product state must be prepared. If sufficient knowledge of a possible reactant and product state is available, $G_{R\leftrightarrow P}$ could be chosen as simple as a linear interpolation of the related coordinates. The geometric model should be continuous, such that each configuration of $G_{R\leftrightarrow P}$ may be specified by a single interpolation variable s.

The search for a trajectory which connects the reactant and the product state regimes is based on selecting putative intermediates $G_{R\leftrightarrow P}(s)$ and assigning random velocities. The velocities should be generated in a way that the resulting trajectory belongs to the desired simulation ensemble. This may be critical in the microcanonical ensemble, which implies $E_{\mathrm{kin}} = E_{\mathrm{tot}} - E_{\mathrm{pot}}$. If $G_{R\leftrightarrow P}(s)$ represents a very unfavorable configuration, the potential energy might be larger than the desired total energy. For starting from such classically forbidden points in phase space, one may however choose $E_{\mathrm{kin}} = 0$ and start TPS at a somewhat larger total energy. In the course of TPS iterations one may then gradually decrease E_{tot} to the desired value.

Let us assume $s = 0$ in the reactant state regime and $s = 1$ for the product state region and first investigate the time propagation of the two configurations $G_{R\leftrightarrow P}(s=0)$ and $G_{R\leftrightarrow P}(s=1)$. When starting a molecular dynamics simulation from a configuration close to the stable reactant or product regime, the resulting trajectory typically evolves to the nearest minimum of the free energy landscape. As this applies to both directions of time propagation, the related pathways lead from reactant to reactant states or from product to product states, respectively. However, by starting from a putative intermediate with $0 < s < 1$ a reactive event, i.e., a trajectory going from reactants to products or vice versa may be found. For this s needs to be chosen sufficiently close to the intersection of $G_{R\leftrightarrow P}(s)$ and the transition state ensemble. By means of an interval bisection procedure such a value for s is usually found within a few iterations.[13,22−27] It is useful to prepare several initial trajectories from various geometric models $G_{R\leftrightarrow P}(s)$, $G'_{R\leftrightarrow P}(s)$, etc., which should differ considerably from each other (Scheme 3). This allows starting TPS in different regimes of trajectory space—ideally at different mechanistic routes—to check pathway convergence as discussed in section 2.3. Examples for this approach are described in detail in refs 17 and 22−25.

It should be stated that a linear interpolation of *all* atomic positions of the reactant and product state configurations may lead to unphysical intersections. To avoid this problem one might reduce the interpolation to a few degrees of freedom like one or two characteristic bond lengths. However, in many cases the system under consideration is too complex to formulate a geometric model from intuition. For example this applies to crystal nucleation from solution, in which only limited knowledge of the explicit arrangement of the reaction products is available. The geometric model may then be prepared from a molecular dynamics run, in which artificial driving forces are applied to enhance the reaction process. This may be incorporated by elevated temperature, pressure, or other thermodynamic driving such as manipulated chemical potentials. The latter approach was used in our recent study of NaCl aggregation from aqueous solution.[13] Therein the van der Waals parameters for the ion−water interactions were changed to lower the solubility of the ions. From this artificial crystallization trajectory configurations were cut and

Scheme 3. Sampling of Transition Pathways Starting from an Intermediate Generated from Geometric Modeling (Dashed Curves)[a]



[a] While trajectory 1) might cross the transition state ensemble via a rather unfavorable configuration, subsequent shooting moves 2), 3), etc. cause pathway evolution towards more preferred reaction routes.

considered as putative intermediates $G_{R\leftrightarrow P}(s)$, where s reflects the time at which the snapshot was taken.

3. TPS in Solution Chemistry

3.1. System Complexity. The most commonly used picture of a reaction relies on the existence of a single, well-defined reactant state. The latter is assumed to be connected to a single product state via 'the' transition state. Processes in solution however take place in complex systems of high dimensionality. The reduction of an ensemble of states to a single point in phase space therefore needs to be considered with caution. Indeed, even for one of the most simple reactions in solution, the dissociation of a $Na^+\cdot\cdot Cl^-$ ion pair in aqueous solution, Chandler and co-workers identified a manifold of transition states.[3,9] On the basis of TPS simulations they generated around 1000 trajectories of this reaction. The analysis revealed the complexity of the underlying mechanism and the importance of solvent degrees of freedom for the understanding of the reaction coordinate.[3,9]

In a recent work, we investigated the formation of NaCl aggregates of around 20 ions from an aqueous solution.[13] From the study of this complex process a variety of different ion aggregates was found. In other words, the product state regime reflects a large area in phase space and may clearly not be reduced to a single ionic arrangement. This phenomenon is related to the interplay of the water molecules and the ions. In aqueous solution the polar water molecules may stabilize the ion aggregates by forming $H_2O\cdot\cdot Na^+$ and $HOH\cdot\cdot Cl^-$ bridges (Figure 1), while in the gas phase the configurational manifold of NaCl clusters of comparable size is significantly lower. [33]

3.2. Investigating Reaction Mechanisms and the Transition State Ensemble. The complexity of the simulation systems encountered in solution chemistry makes the investigation of reaction mechanisms difficult yet not entirely impossible. For reactions in solution the solvent usually plays an important role and solvent degrees of freedom hence are part of the reaction coordinate. While the various types of solvent effects are specified in the next section, we shall first focus on more technical aspects for identifying reaction mechanisms and transition states of processes in complex systems in general.

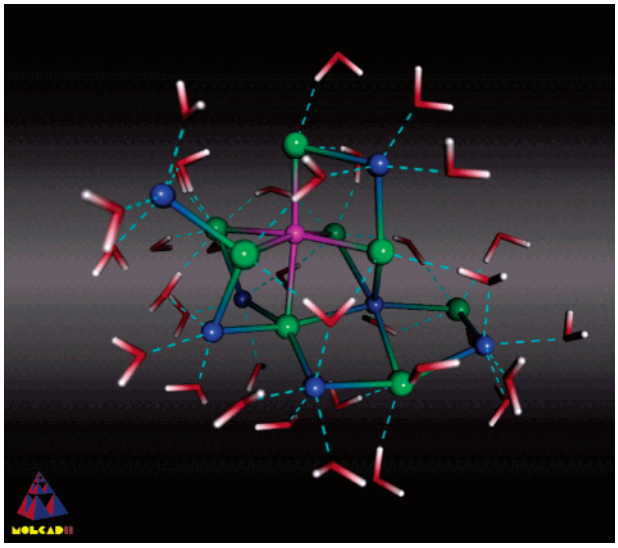


Figure 1. Na_8Cl_8 aggregate in aqueous solution as obtained from TPS molecular dynamics simulations 8. Sodium and chloride ions are colored in blue and green, respectively. The sodium ion of the $Na^+Cl^-_6$ octahedra is highlighted in purple.

When analyzing the NaCl aggregates discussed in ref 13, we identified common features in each of the reaction pathways. For one of the aggregates this is illustrated in Figure 1. Roughly in the center of the aggregate a sodium ion is observed, which exhibits no water molecule in its first coordination sphere. Instead, it is octahedrally coordinated by six chloride ions. While the arrangement of the remaining ions varies considerably, the $Na^+Cl^-_6$ octahedron forms a stable core in the aggregates. This motif of the NaCl crystal structure was found to be a common feature and was therefore proposed as characteristic for the formation of stable aggregates of around 20 Na^+ and Cl^- ions. In more general terms, we investigated the reactive pathways for common features and interpreted them as aspects of the reaction mechanism. This strategy proved quite effective in a series of studies related to reactions in complex systems.[11−13,17]

For the identification of common features in reactive pathways we recommend to also investigate the trajectories of failed attempts generated in the course of TPS iterations. Each of these trajectories was derived from small variations of a successful reaction pathway. The failed attempts therefore often represent pathways, in which the reaction almost took place, but at least one important contribution to the reaction mechanism was missing. Comparing such pathways with the trajectories of successful reaction attempts may help a lot in finding detailed information of the reaction mechanism. While the main characteristics of the reaction mechanism are usually easy to observe from the ensemble

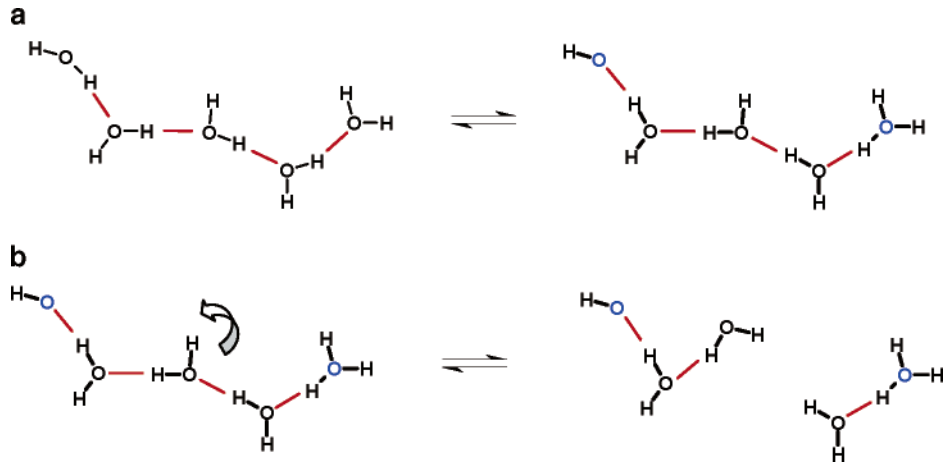


Figure 2. a. First step of the water autodissociation as observed from TPS Car-Parrinello molecular dynamics simulations 4. The formation of an $H_3O^+ \cdot\cdot OH^-$ contact ion pair is avoided by multiple proton-transfer steps resulting in charge separation over several water molecules. b. The hydrogen bonded chain of water molecules, which allowed fast OH^- and H^+ transport, is broken. This prevents the fast recombination of the separated charges.

of successful trajectories, contrasting true reaction pathways to failed attempts is particularly suitable for identifying the fine details.

The approaches described above should help understanding reaction mechanisms at least from a qualitative point of view. Some of the 'common features' characterizing the reaction mechanism may actually be variables that can be clearly defined and hence used to construct the reaction coordinate. For examples, this holds for bond distances and angles. However, for more complex features of the reaction mechanism the determination of explicit variables is typically much more complicated. A very elegant way of performing a reaction coordinate analysis within an automated scheme was recently presented by Ma and Dinner.[16] In this work a large set of variables is related to the committor analysis of the reaction by means of artificial intelligence. Performing neural network calculations Ma and Dinner succeeded to isolate a small number of relevant variables which were demonstrated to be sufficient for describing the $C_{7eq} \rightarrow \alpha_R$ isomerization of the alanine dipeptide in aqueous solution.[16]

An important contribution to a deeper understanding of reaction mechanisms may be provided from exploring the ensemble of transition states. The underlying committor analysis is described in detail in refs 1, 2, 28, and 29 and shall be summarized only briefly here. Following the definition of Du et al. the transition states represent configurations in real space, which—after assigning random velocities—will evolve to either the reactant of product state regime at equal probability.[34] The transition state analysis may hence be accomplished by the following scheme: for each reaction pathway a series of snapshots is chosen. For each of these snapshots $\{r_i\}_{i=1..Natoms}$ a number of may be 100 phase points $\{r_i,v_i\}_{i=1..Natoms}$ is prepared by combining the atomic positions r_i with different sets j of random velocities $v_i(j)$ generated in accordance to the desired simulation ensemble. Then the time propagation of each configuration $\{r_i,v_i(j)\}$ is investigated from molecular dynamics simulations. The different velocity sets j provide a statistical estimate of the probability p_R of $\{r_i\}$ to evolve to the reactant state regime. The manifold of transition states comprises all configurations $\{r_i\}$ with $p_R(\{r_i\}) = 0.5$. It should be noted that a single reaction pathway may cross the transition state ensemble several times before connecting the reactant and the product state regimes.

3.3. The Solvent Effect. Depending on the simulation system the solvent may be involved in different ways in the reaction process. The most direct role the solvent can play is that of a *possible reactant*. A prominent example for this issue is given by proton transfer reactions in aqueous solution. Therein the water molecules may act as proton donors and acceptors. Moreover, protons may be transported along the hydrogen-bonded network of water molecules via a Grotthuss type mechanism. The importance of this phenomenon for the autodissociation of water was recently demonstrated from TPS Car-Parrinello molecular dynamics simulations.[6] The direct formation of an $H_3O^+ \cdot\cdot HO^-$ contact ion pair is disfavored by the strong tendency of recombining the separated charges.[35] Instead, the dissociation involves multiple proton-transfer steps resulting in oxonium and hydroxide ions, which are separated by several coordination spheres. The overall reaction hence reads $H_2O + n \cdot H_2O + H_2O \rightleftharpoons H_3O^+ + n \cdot H_2O + OH^-$ (with $n \geq 3$ and $n=3$ in Figure 2a). To stabilize the right-hand side of Figure 2a the hydrogen bonded chain connecting the separated charges must be broken (Figure 2b). As a consequence, water dissociation not only implies the formation of a specific solvent arrangement to favor the forward reaction but also requires the dissociation of the assisting chain of hydrogen bridged water molecules to avoid back-reaction. An analogous picture was recently observed for the rate-determining step in acid-catalyzed amide hydrolysis in aqueous solution.[11] Therein a water molecule performs a nucleophilic attack on the amide bond by adding an OH^- group to the amide and transferring a proton to the solvent. Contrasting reactive trajectories and failed attempts, we found that the formation of stable reaction products requires further proton-transfer steps leading to H^+ migration to the aqueous solvent. This process occurs in the same way as observed in the water dissociation reaction: The proton migration is assisted by a hydrogen-bonded chain of several water molecules, which must be disconnected after

the reaction took place in order to avoid immediate back-reaction. This phenomenon might play an important role in acid/base reactions in aqueous solutions in general and clearly should be considered in mechanistic studies of such processes.

A less obvious yet important solvent effect observed for reactions in solution is related to *different energetic and/or entropic favoring* of the reactant, transition, or product state ensemble. For polar solvents, this phenomenon mainly accounts for the Coulomb interaction of the reacting molecules and the embedding media. This type of solvent effect is often modeled by an electrostatic continuum approach. However, for charge transfer reactions such as the autodissociation of water described above, the fluctuations of the electric field induced by the solvent are of key importance. Spontaneously formed solvent arrangements may trigger the reaction by lowering the reaction barrier or even fully biasing the reacting system in favor of a product state.

While simulation studies based on static approaches can only identify correlations of specific solvent arrangements and the reaction process, TPS molecular dynamics simulations allow the investigation of a time-resolved picture. Impressive examples for such studies were presented by Chandler and co-workers, who investigated the flux of water molecules during the dissociation of NaCl ion pairs[3,9] and the role of solvent fluctuations in the water autoionization process.[6]

A purely kinetic aspect of the solvent effect is reflected by its role as a *heat bath*. Reactions, which require the crossing of energetically disfavored intermediates, imply the accumulation of sufficient kinetic energy to allow the system to overcome the energy barrier. Before the reaction takes place, the system must therefore 'focus' kinetic energy to the reaction coordinate degree of freedom. This usually occurs at the cost of perpendicular modes. This effect may be illustrated from our recent study of helium insertion into a C_{60} buckyball.[12] Instead of a polar solvent, which would predominantly interact via Coulomb forces, this simulation model comprised of a box of 1000 helium atoms mimicking an autoclave scenario. Prior to the penetration of the C_{60} by a helium atom, we observed a series of collisions in the gas atmosphere. These collisions occur in such a way that the momentum of one of the helium atoms increases at the cost of the kinetic energy of the other. In the successful reaction attempts, this process accumulated sufficient kinetic energy on a single helium atom and directed its momentum toward the buckyball molecule, such that the helium crossed the insertion barrier.

The helium atom, which penetrates the C_{60}, was observed to use almost all of its kinetic energy for overcoming the potential energy barrier. However, after crossing the transition state, the helium atom regains kinetic energy when approaching the product state. This kinetic energy is sufficiently high to allow recrossing of the potential energy barrier and must therefore be dissipated to other degrees of freedom to avoid immediate back-reaction. Indeed, some of the failed reaction attempts exhibited a helium insertion, followed by reflection at the inner wall of the buckyball molecule and expulsion in opposite direction of the insertion route.

4. Conclusions

We reviewed a series of molecular dynamics studies of reactions in solution using the TPS approach. Typically, reactions in solution are complex, and their investigation may particularly benefit from the advantages of the TPS simulation scheme. Therein the mechanistic study can be based on a manifold of reaction pathways and a series of trajectories related to failed attempts. Contrasting both classes of pathways offers very profound insights into the reaction dynamics including the role of the solvent molecules. The solvent effect may be rated to several phenomena including catalytic functions, energetic, and/or entropic favoring and the role of a heat bath.

TPS may be combined to all variations of molecular dynamics simulations, including classical,[3−5,7,9,13,15,16] mixed quantum/classical,[10,12,17] and ab initio[6,8,11,14] approaches. The study of reactions in solution typically requires including a large number of solvent molecules to the simulation model and therefore implies considerable computational efforts. A series of tricks of the trade collected from several recent studies of rare events in complex systems is summarized and discussed in detail.

Acknowledgment. The author wishes to thank Rüdiger Kniep, Stefano Leoni, Claire Loison, and Francesco Mercuri for many fruitful discussions. Further acknowledgments are dedicated to the participants of the CECAM workshops "Simulation of rare events: The reaction coordinate problem in complex systems" and "Conformational dynamics in complex systems", whose questions inspired the preparation of this review.

References

(1) Bolhuis, P. G.; Dellago, C.; Chandler, D. *Faraday Discuss.* **1998**, *110*, 421.

(2) Dellago, C.; Bolhuis, P. G.; Csajka, F. S.; Chandler, D. *J. Chem. Phys.* **1998**, *108*, 1964.

(3) Geissler, P. L.; Dellago, C.; Chandler, D. *J. Phys. Chem. B* **1999**, *103*, 3706.

(4) Marti, J.; Csajka, F. S. *J. Chem. Phys.* **2000**, *113*, 1154.

(5) Marti, J.; Csajka, F. S.; Chandler, D. *Chem. Phys. Lett.* **2000**, *328*, 169.

(6) Geissler, P. L.; Dellago, C.; Chandler, D.; Hutter, J.; Parrinello, M. *Science* **2001**, *291*, 2121.

(7) Marti, J. *Mol. Simul.* **2001**, *27*, 169.

(8) Ensing, B.; Baerends, E. J. *J. Phys. Chem. A* **2002**, *106*, 7902.

(9) McCormick, T. A.; Chandler, D. *J. Phys. Chem. B* **2003**, *107*, 2796.

(10) Chu, J. W.; Brooks, B. R.; Trout, B. L. *J. Am. Chem. Soc.* **2004**, *126*, 16601.

(11) Zahn, D. *Chem. Phys.* **2004**, *300*, 79.

(12) Zahn, D.; Seifert, G. *J. Phys. Chem. B* **2004**, *108*, 16495.

(13) Zahn, D. *Phys. Rev. Lett.* **2004**, *92*, 40801.

(14) Lo, C. S.; Radhakrishnan, R.; Trout, B. L. *Catal. Today* **2005**, *105*, 93.

(15) Snee, P. T.; Shanoski, J.; Harris, C. B. *J. Am. Chem. Soc.* **2005**, *127*, 1286.

(16) Ma, A.; Dinner, A. R. *J. Phys. Chem. B* **2005**, *109*, 6769.

(17) Zahn, D. *J. Chem. Phys.* **2005**, *123*, 44104.

(18) ten Wolde, P. R.; Chandler, D. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 6539.

(19) Bolhuis, P. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 12129.

(20) Hagen, M. F.; Dinner, A.; Chandler, D.; Chakraborty, A. K. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13922.

(21) Pan, A.; Chandler, D. *J. Phys. Chem. B* **2004**, *108*, 19681.

(22) Zahn, D.; Leoni, S. *Phys. Rev. Lett.* **2004**, *92*, 250201.

(23) Zahn, D. *Phys. Rev. Lett.* **2004**, *93*, 227801.

(24) Leoni, S.; Zahn, D. *Z. Kristallogr.* **2004**, *219*, 339.

(25) Zahn, D.; Leoni, S. *Z. Kristallogr.* **2004**, *219*, 345.

(26) Zahn, D. *J. Solid State Chem.* **2004**, *177*, 3590.

(27) Leoni, S.; Zahn, D. *Z. Anorg. Allg. Chem.* **2004**, *630*, 1738.

(28) Dellago, C.; Bolhuis, P. G.; Geissler, P. L. *Adv. Chem. Phys.* **2002**, 123.

(29) Bolhuis, P. G.; Dellago, C.; Geissler, P. L.; Chandler, D. *Annu. Rev. Phys. Chem.* **2002**, *54*, 20.

(30) van Erp, T. S.; Moroni, D.; Bolhuis, P. G. *J. Chem. Phys.* **2003**, *118*, 7762.

(31) Moroni, D.; Bolhuis, P. G.; van Erp, T. S. *J. Chem. Phys.* **2004**, *120*, 4055.

(32) Wang, F.; Landau, D. P. *Phys. Rev. Lett.* **2001**, *86*, 2050.

(33) Kawska, A.; Hochrein, O.; Brickmann, J.; Zahn, D. *Z. Anorg. Allg. Chem.* **2005**, *631*, 1172.

(34) Du, R.; Pande, V. S.; Grosberg, A. Y.; Tanaka, T.; Shakhnovich, E. S. *J. Chem. Phys.* **1998**, *108*, 334.

(35) Trout, B. L.; Parrinello, M. *Chem. Phys. Lett.* **1998**, *288*, 343.

CT0501755

JCTC Journal of Chemical Theory and Computation

Investigation of Salt Bridge Stability in a Generalized Born Solvent Model

Raphaël Geney,[†] Melinda Layten,[§] Roberto Gomperts,[#] Viktor Hornak,[‡] and Carlos Simmerling*[,†,‡,§]

Department of Chemistry, Graduate Program in Molecular and Cellular Biology, and Center for Structural Biology, Stony Brook University, Stony Brook, New York 11794-3400, and Silicon Graphics Inc., Applications Engineering Group, Hudson, Massachusetts 01749

Received July 29, 2005

Abstract: Potentials of mean force (PMFs) of salt bridge formation between oppositely charged amino acid side chains were calculated both in explicit solvent and in a Generalized Born (GB) continuum solvent model to quantify the potential overstabilization of side chain ion pairs in GB relative to explicit solvation. These show that salt bridges are too stable by as much as 3−4 kcal/mol in the GB solvent models that we tested, consistent with previously reported observations of significantly different structural ensembles in GB models and explicit solvent for proteins containing ionizable groups. We thus investigated a simple empirical correction, wherein the intrinsic GB radii of hydrogen atoms bound to charged nitrogen atoms are reduced, effectively increasing the desolvation penalty of the positively charged groups. The thermodynamics of salt bridge formation were considerably improved, as exemplified by the close match of the corrected GB PMF to the reference explicit solvent PMF, and more significantly by our ability to closely reproduce the experimental temperature melting profile of the TC5b Trp-cage miniprotein, which is otherwise highly distorted by prevalent non-native salt bridges when using standard GB parameters.

Introduction

One of the greatest challenges in the application of computation techniques to biological systems is the accurate determination of protein and RNA three-dimensional structures. The native structure of proteins is maintained at the edge of thermodynamic stability, the free energy of unfolding being in the range of a few kcal/mol. A dominant contributor to stability is the hydrophobic effect, but other important stabilizing factors include van der Waals interactions, hydrogen bonds, and electrostatic interactions, notably salt bridges.[1−3] However, with the stability of salt bridges, i.e., the net balance of favorable Coulombic interactions between opposite charges and their costly desolvation as well as the extent of their involvement in native state stabilization remain ambiguous.[4−9] Nevertheless, salt bridges have been linked to the thermal stability of hyperthermophilic proteins.[1,10−15]

Molecular simulations have proven to be valuable tools for probing the various interactions that define the protein native state and characterize possible folding pathways toward it.[16] Recently, continuum solvent simulations[17−22] have become popular alternatives to their more computationally demanding explicit solvent counterparts, as their lack of solvent friction increases conformational transition rates significantly,[23−31] allowing for faster sampling of the configurational space. Furthermore, because continuum solvent models implicitly average over the water and counterion distributions, this averaging does not need to be done by

* Corresponding author phone: (631)632-1336; e-mail: carlos.simmerling@stonybrook.edu.
† Department of Chemistry, Stony Brook University.
‡ Center for Structural Biology, Stony Brook University.
§ Graduate Program in Molecular and Cellular Biology, Stony Brook University.
# Silicon Graphics Inc.

10.1021/ct050183l CCC: $33.50
Published on Web 11/11/2005

the simulation itself which leads to considerable simplification when calculating thermodynamic properties.[32] Last, some macroscopic solvent properties, such as dielectric effects, are difficult to reproduce accurately with explicit solvation.[33] The ability to build these into implicit solvent descriptions may actually give them some advantage for certain kinds of simulations.

Due to its computational efficiency, the Generalized Born (GB) implicit solvent model[34−36] has become a popular choice to accelerate molecular dynamics (MD) simulations and to study large scale conformational transitions. However, this model lacks structural water features and has been reported to yield higher fluctuations than explicit solvent simulations.[37] To some extent, this might be a consequence of the improved conformational sampling, which lets the simulation more quickly find non-native structures that are energetically favored by the particular force field. But it also seems likely that current GB models do not have as good a balance between protein−protein and protein−solvent interactions as do the more widely tested explicit solvent models. More particularly, we[38] and others[39−42] have observed that salt bridges were frequently too stable in the GB implicit water model, causing salt bridged conformations to be oversampled in MD simulations, thus altering the thermodynamics and kinetics of folding for small peptides. A clear illustration was given by Zhou and Berne,[40] who sampled the C-terminal β-hairpin of protein G (GB1) with both a surface-GB (SGB)[43] continuum model and explicit solvent using a replica-exchange molecular dynamics (REMD)[44] protocol. The lowest free energy state with SGB was significantly different from the lowest free energy state in explicit solvent, with incorrect salt bridges formed at the core of the peptide, in place of hydrophobic contacts. Zhou extended this study on GB1 by examining several force field-GB model combinations,[39] with all GB models showing erroneous salt-bridges. Nevertheless, as the MD simulation community envisions characterizing entire folding landscapes and pathways, implicit solvent models such as GB could be beneficial in supplementing the more slowly converging explicit models but should be devoid of structural bias in order to maintain comparable levels of accuracy.

In this study, the Potential of Mean Force (PMF) of salt bridge formation is calculated for two residues in a solvated protein environment. Masunov and Lazaridis[45] performed similar calculations on isolated side-chain pairs in coplanar monodirectional approaches and concluded that CHARMM GB[46] matches the explicit solvent contact minimum energy to within 1 kcal/mol for both the $Arg^+\cdots Glu^-$ and $Lys^+\cdots Glu^-$ pairs. In our case, comparing salt bridge PMFs obtained either in the GB^{HCT} model[47−50] of AMBER or TIP3P explicit water[51] confirms the excessive strength of salt bridges in this GB model and offers a way to assess its parametrization. A simple empirical change in the assignment of dielectric radii for hydrogen atoms of charged protein groups is investigated and shown to significantly improve the GB PMF of our test salt bridge system. This parameter change is further examined on a range of control systems by comparison to explicit solvent and experimental data.

Methods

All calculations were performed using the AMBER suite of programs,[52] versions 7 and 8, with the ff99 force field[53] modified to improve agreement with ab initio relative energies of alanine tetrapeptide conformations (frcmod.mod_phipsi.1).[38,54] The Trp-cage simulations employed an optimized version of this force field, refit to also reproduce ab initio relative energies of the Gly tetrapepide.[55]

Bonds involving hydrogen atoms were constrained using the SHAKE algorithm,[56] and a 2 fs integration time step was adopted. Explicit solvent simulations were performed with the TIP3P water model,[51] widely popular for its computational simplicity and near-experimental bulk permittivity.[33,57−59] The Fab 17/9 (PDB ID:1HIL[60]) H3 loop fragment and the small helical peptides were placed in truncated octahedral boxes with respectively 5 or 6 Å minimum buffer clearance from the solute.

The Particle Mesh Ewald[61−64] (PME) treatment of long-range electrostatics was used with a direct space cutoff of 8 Å in constant pressure simulations at 1 bar. Implicit solvent runs employed the GB^{HCT} model[47,49,50] with modified Bondi radii[65] and no cutoff. Radii reductions in GB^{HCT} were further applied to hydrogen atoms bonded to nitrogens of N2 and N3 AMBER types,[66] as in Arg, Lys, and charged N-terminal groups.

In the PMF calculations, the varied reaction coordinate was the distance between the carboxyl carbon of the acidic side chain (Cγ in Asp, Cδ in Glu) and either Nζ of Lys or Cζ of Arg, the geometric center of the ionized guanido group. In all cases, the backbone atoms were positionally restrained with sufficient force to prevent significant conformational changes (1 kcal/mol·Å^2 force constant for the Fab 17/9 H3 loop, 10 kcal/mol·Å^2 for the test helical peptides). All PMFs were calculated using Umbrella Sampling (US)[67] with the reaction coordinate constrained to a narrow range by application of a harmonic biasing potential $V(r) = k_{umb}(r - r_0)^2$. US windows were centered every 0.5 Å of the coordinate range (3−11.5 Å) and a 1 ns MD run was performed for each. Umbrella potentials with $k_{umb} = 10$ kcal/Å^2 were applied in all windows, to enforce continuous sampling of high-energy regions. The biased frequency distributions were converted to free energies using the WHAM method,[68] as implemented by Roux.[69] Additional windows were placed at 3.25, 3.75, 4.25, 4.75, 5.25, 5.75 Å for TIP3P simulations, to improve sampling of the barrier region. Data from the first 200 ps of each window were discarded.

To extensively explore the conformational space of the TC5b miniprotein (NLYIQWLKDGGPSSGRPPPS) in the GB^{HCT} solvent model, we employed replica-exchange molecular dynamics simulations (REMD[44,70]) as implemented in AMBER 8. TC5b was modeled in its zwitterionic form, with ionizable residues in their expected ionization state at pH = 7, for a total of 304 atoms. The 267−715.7 K temperature range was covered using 16 replicas (267.0, 285.1, 304.5, 325.2, 347.3, 370.9, 396.1, 423.0, 451.8, 482.4, 515.2, 550.2, 587.6, 627.5, 670.2, 715.7 K), resulting in average exchange acceptance probabilities in the 22−32% range. Exchanges were attempted, and replica conformations were recorded every 500 MD steps (1 ps).

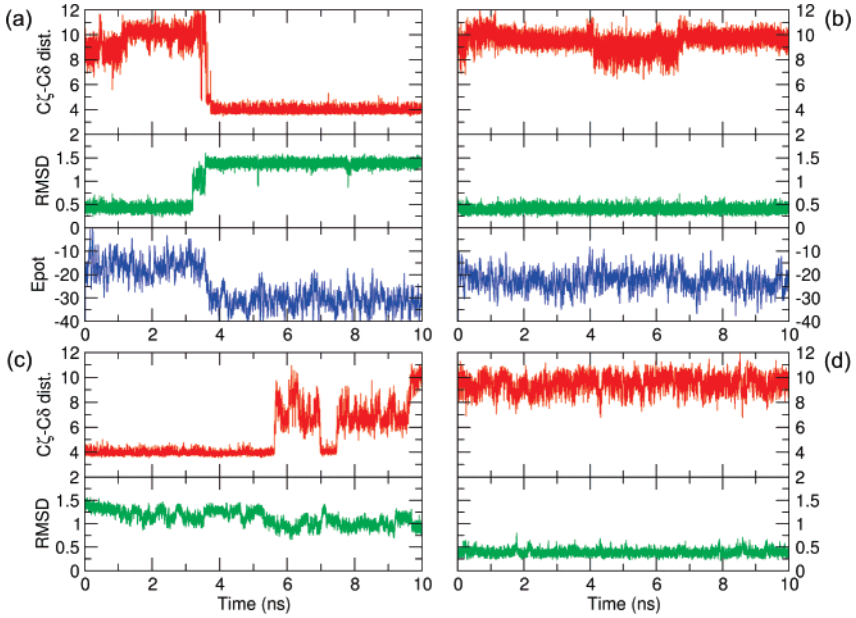


Figure 1. Root-mean-square deviation (RMSD, Å) of Fab 17/9 H3 loop backbone heavy atoms and salt bridge distance (Arg^{97} Cζ−Glu^{100} Cδ, Å) as a function of simulation time, in different conditions at 300 K: (a) GB^{HCT} from native, (b) GB^{HCT} with uncharged Arg^{97} and Glu^{100} side chains from native, (c) TIP3P explicit solvent simulation from salt bridge conformation, and (d) TIP3P from native. The RMSD fit to the X-ray conformation is performed over the restrained, nonloop atoms of the fragment. Relative potential energy values, window averaged over 25 ps, are also reported for implicit solvent simulations in kcal/mol. The backbone transition and concomitant salt bridge formation in GB^{HCT}, not observed with neutralized side chains or explicit solvent, induce a 14 kcal/mol reduction in potential energy.

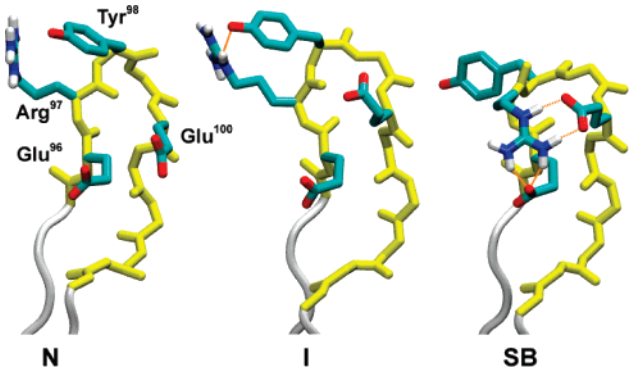


Figure 2. Fab 17/9 H3 loop in native conformation (N), transient intermediate state with inverted Tyr^{98} ψ and Asp^{99} ϕ dihedral angles (I), and stable salt bridged conformation with bidentate H-bond (SB), taken from a standard GB^{HCT} simulation. Loop backbone heavy atoms are colored yellow, while selected side chains are colored by element. H-bonds are indicated by dashed orange lines.

After a 9 ns thermal equilibration period, data were accumulated for 50 ns for each temperature. In GB^{HCT} with standard H^{N+} radii (1.3 Å), the salt bridge strength hampered sampling, and REMD runs were extended to 92 ns in an effort to achieve reasonable convergence.

The Berendsen temperature coupling scheme[71] was applied with a 0.1 ps heat bath coupling constant for all replicas (1 ps for non-REMD simulations). To test the influence of this particular thermostat, the GB^{HCT} PMF profile shown in Figure 4 was recalculated using Langevin dynamics and collision frequency of 1 ps^{-1}. The profile remained essentially unchanged, with a maximum deviation of ∼0.5 kcal/mol from that obtained using Berendsen coupling.

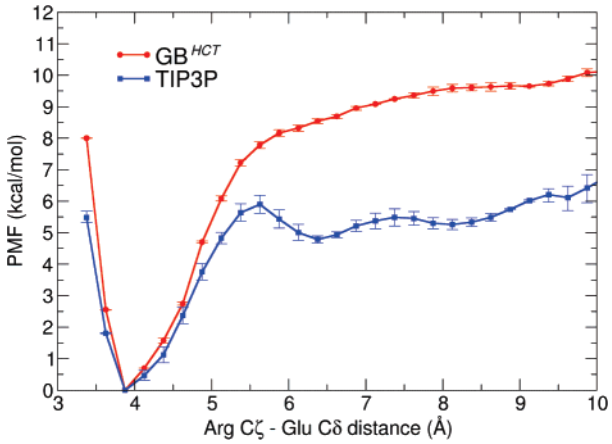


Figure 3. Fab 17/9 antibody Arg^{97}···Glu^{100} ion pair PMF as a function of the intercharged groups distance (Arg^{97} Cζ−Glu^{100} Cδ), at 300 K. The GB^{HCT} PMF overestimates the contact ion pair stability by as much as ∼4 kcal/mol. The loop backbone conformation is restrained in the SB state (Figure 2). Error bars on both curves estimate the sampling error and were derived by separately considering the first or the second half of the data set.

Lower bound estimates of the sampling uncertainty and convergence of our simulation protocols were derived by splitting data sets in half and comparing individual half-length averages to the full-length values.

Results and Discussion

Unstable Behavior of the 17/9 Anti-Influenza Fab H3 Loop in GB Simulations. In the course of our research on loop structure modeling,[72,73] our attention focused on the H3

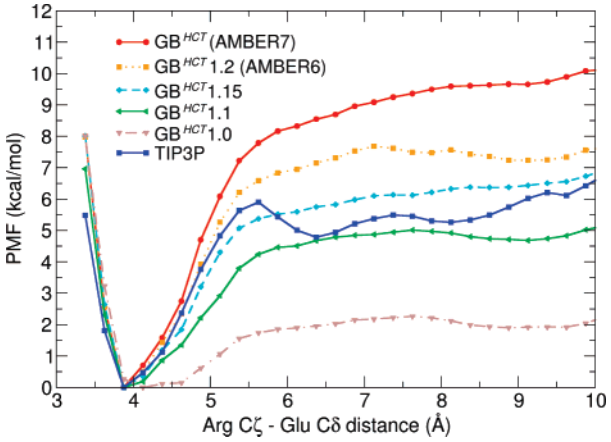


Figure 4. Potentials of mean force for the Fab 17/9 H3 loop $Arg^{97}\cdots Glu^{100}$ ion pair in different solvent models, at 300 K. The distance coordinate is measured between Cζ of Arg^{97} and Cδ of Glu^{100}.

CDR loop of the 17/9 anti-influenza antibody.[60] In preliminary GB^{HCT} MD simulations, its experimentally determined native structure appeared highly unstable, with backbone transitions of nearly 1 Å RMSD magnitude occurring within a few nanoseconds (Figure 1a). This seems unlikely to be attributable to a poorly refined experimental structure, since the 1HIL structure was solved at a reasonable 2.0 Å resolution, and the H3 loop under study is rather well defined, with maximum backbone and side-chain atomic B-factors of 22.08 and 44.82 $Å^2$, respectively. Nevertheless, this loop incorporates intrinsic flexibility as revealed by crystallography studies that indicate substantial loop rearrangement occurs upon binding to a nonapeptide antigen.[60]

The H3 loop and its surroundings incorporate numerous charged residues, and during simulation, the flexible Arg^{97} side chain (the Kabat antibody sequence numbering convention[74] is followed throughout this paper) associates with Glu^{100}, thereby irreversibly shifting the backbone in a bent non-native conformation (Figure 2). This transformation happens through an intermediate where Tyr^{98} and Asp^{99}, at the tip of the loop, have simultaneously undergone backbone conformational transitions (Figure 2), yet maintaining the hydrogen bond observed between the Arg^{97} and Tyr^{98} side chains in the native state. Rapidly following, the last step of this transformation is the conversion of Arg^{97} from a polyproline II to a left-handed α-helix conformation, simultaneous to salt bridge formation (Figure 2). This last step generates a ∼14 kcal/mol drop in potential energy, which effectively locks the loop in the non-native conformation.

Artificially neutralizing both the Arg^{97} and Glu^{100} side chains prevented this behavior (Figure 1b), clearly suggesting the electrostatic nature of the phenomenon, and in particular an imbalance between GB desolvation energy and Coulombic attraction. This control run also evidenced that backbone parameters alone are not responsible for the observed conformational transition.

Even more intriguing was the fact that a TIP3P/PME explicit solvent simulation initiated from the salt-bridged structure saw opening of the $Arg^{97}\cdots Glu^{100}$ ion pair (Figure 1c). However, this was not accompanied by rearrangement of the backbone back to the X-ray conformation during the 10 ns, a process assumed to be slow in explicit solvent. The native loop conformation was also stable throughout a 10 ns TIP3P/PME run started from the X-ray structure, with no salt bridge formation observed (Figure 1d).

PMFs as Measures of GB Deviation from Explicit Solvent Behavior. Owing to the difficulty of directly comparing simulations with experimental salt bridge stability data, mostly of mutational origin, explicit water simulations were chosen as our reference for evaluating the PMF profile of the Fab 17/9 H3 loop $Arg^{97}\cdots Glu^{100}$ salt bridge in GB. As no computationally tractable model—especially not the rigid nonpolarizable model used here—is presently able to correctly reproduce all experimental properties of water,[33] we do not expect to accurately reproduce experimental ion pair behavior. However, the inclusion of solvent molecularity provides a significantly less crude approach than the ad hoc GB model and is used here for consistency with previously published ion pair solvation studies.[45,75−80]

PMFs were obtained in GB^{HCT} and TIP3P explicit solvent using umbrella sampling[67] along the interside-chain distance, with the loop backbone restrained in the SB conformation (Figure 2). The resulting TIP3P profile (Figure 3) consists of a series of well-defined minima: the contact ion pair (CIP) at 3.9 Å, corresponding to the free energy minimum, is accompanied by two solvent-separated ion pair (SSIP) minima at 6.4 Å and 8.2 Å, corresponding to the insertion of one or two TIP3P molecules in the interside-chain volume.[76,81] Qualitatively, the overall shape of the PMF is in good agreement with that reported by Lazaridis for an isolated $Arg^+\cdots Glu^-$ pair in a coplanar monodirectional approach.[45] Quantitatively, however, our method yields a barrier height of 6 kcal/mol for going from the CIP to the first SSIP in TIP3P, while the PMF they reported for the isolated $Arg^+\cdots Glu^-$ ion pair in the coplanar, double H-bonded approach presents a 7.7 kcal/mol barrier to escape the contact minimum.[45] This slight difference is readily justified by the different solvent exposure levels, approach geometries (cf. Figure 4a,b of ref 45), and presence of a very polar environment around the Fab 17/9 ion pair, with the possibility for Arg^{97} to also interact with Glu^{96}. Gruia et al. similarly calculated the potential of mean force of the $Arg^{105}\cdot\cdot\cdot Glu^{135}$ salt bridge on the surface of truncated Staphylococcal nuclease (SnaseΔ), after observing in explicit water molecular dynamics simulations that breaking this salt bridge was the rate limiting step of the early unfolding transition.[82,83] Using umbrella sampling, they measured a ∼7 kcal/mol transition barrier height for breaking the contact minimum of the two charged side chains.

In contrast to the TIP3P profile, the GB^{HCT} salt bridge PMF shows no depiction of the various SSIPs and grossly overestimates the TIP3P CIP-SSIP energy difference by 3.8 kcal/mol. The activation energy barrier to breaking the salt bridge is also overestimated by almost 2 kcal/mol, providing clear direct evidence for our hypothesis that salt bridges were too stable in this GB model.

The manifestly insufficient desolvation penalty experienced by the salt bridge in GB^{HCT} prompted us to reexamine

the parametrization of this GB solvent model and in particular its handling of cationic protein side chains.

GB Model Parametrization and Rationale for Reduced H^{N+} Radii. The original Born model computes the electrostatic reversible work required to move a charged sphere from a vacuum environment into a continuous high dielectric region. The result is proportional to the square of the charge and inversely proportional to the size of the ion.[84] These ideas were extended to the case of nonspherical solutes in the generalized Born theory,[34,35] which evaluates the electrostatic component of the solvation free energy in the following way:

$$\Delta G_{\mathrm{elec}} = -\frac{1}{2}\sum_i\sum_j \frac{q_i q_j}{f_{\mathrm{GB}}}\left(1 - \frac{1}{\epsilon_{\mathrm{out}}}\right) \quad (1)$$

f_{GB} is designed to interpolate between an effective Born radius α_i at short interatomic distance r_{ij}, and r_{ij} itself at long distances. Various functional forms are possible for f_{GB}, but AMBER employs the analytically differentiable one originally proposed by Still et al.:[35]

$$f_{\mathrm{GB}}(r_{ij},\alpha_i,\alpha_j) = \left[r_{ij}^{\,2} + \alpha_i\alpha_j \exp\left(-\frac{r_{ij}^{\,2}}{4\alpha_i\alpha_j}\right)\right]^{1/2} \quad (2)$$

The effective Born radius α_i corresponds to the radius that would return the electrostatic energy of the system using the original Born equation if all atoms $j \neq i$ in the solute were uncharged. Therefore, α_i reflects the degree of burial of atomic charge q_i from the solute−solvent dielectric boundary. The computation of effective radii in the particular AMBER GB model discussed here (GB^{HCT})[49,50] follows the pairwise descreening approximation (PDA) of Hawkins et al.,[47,48] wherein the molecule is described as a set of atomic spheres of radii ρ_i (eq 3). The corresponding volume integrals can be calculated analytically even when spheres i and j overlap, following eq 13 in ref 47. An additional atom-dependent screening parameter S_i is required in order to avoid overcounting overlap volume between two or more neighboring spheres j, leading to eq 4 which relates all atomic input parameters

$$\alpha_i^{-1} = \rho_i^{-1} - \frac{1}{4\pi}\sum_{j\neq i}\int_{\mathrm{sphere}_j}\frac{1}{r^4}dV \quad (3)$$

with

$$\rho_i = S_i(R_i + b_{\mathrm{offset}}) \quad (4)$$

Although many combinations of S_i, R_i, and b_{offset} could be used, the GB^{HCT} model of AMBER employs screening parameters from the TINKER molecular modeling package[85] and Bondi radii[65] slightly modified for hydrogen atoms, to reflect their bonding environment (Table 1).[49,50] The original b_{offset} value of 0.09 Å, suggested by Still et al.[35] is employed for GB simulations of proteins in AMBER.

Reparametrization of GB^{HCT} for Improved Handling of Ionic Interactions. To correct the stability of the native Fab 17/9 H3 loop conformation, we reasoned that smaller effective GB radii for atoms involved in the salt bridge would increase their desolvation penalty, thus balancing an otherwise dominating Coulombic attraction. Intuitively, formally charged nitrogens can be seen as having increased electronegativity relative to uncharged nitrogens. This should translate in hydrogen atoms bonded to them (H^{N+} atoms) being assigned smaller dielectric radii than H^N atoms, following the suggestion by Tsui and Case that hydrogen GB radii should decrease with increasing electronegativity of their bonding partner.[49] This reasoning is further substantiated by the lower electron density around H atoms in the ammonium ion, relative to ammonia (14% decrease, based on HF/6-31+G* calculations; data not shown). Radii reductions were applied only to hydrogens bonded to nitrogens of N2 and N3 AMBER types,[66] as found in Arg, Lys, and charged N-terminal residues. His protons were not considered, thus far, as their involvement in ionic pairs is less frequent[86,87] and generally weak.[45] Reducing only the radii of H^{N+} atoms also does not greatly affect the overall protein solvation energy (<8% in our tests), while specifically weakening the ion pair. Interestingly, similar ad hoc corrections have been recently proposed by both the Levy and Honig groups, in which additional dielectric screening is applied to oxygen and nitrogen atoms of formally charged groups either through eq 2[42] or eq 4.[88]

Table 1. Parameter Sets Used in the Various AMBER GB Implementations[a]

atom	R_i, GB^{HCT} AMBER6[49]	R_i, GB^{HCT} AMBER7,8[50]	S_i[85]
H^C	**1.3**	**1.3**	0.85
H^N	**1.2**	**1.3**	0.85
H^O	**0.8**	**0.8**	0.85
H^S	**0.8**	**0.8**	0.85
C	1.7	1.7	0.72
N	1.55	1.55	0.79
O	1.5	1.5	0.85
F	1.5	1.5	0.88
P	1.85	1.85	0.86
S	1.8	1.8	0.96

[a] The superscript on H atoms indicates the heavy atom to which it is bound. The H^N R_i value was increased in AMBER7 to stabilize Watson−Crick hydrogen bonds in a 10-base pair DNA duplex.[50]

The GB^{HCT} salt bridge PMF profiles are very sensitive to the choice of H^{N+} radii applied, as a 0.1 Å decrease in radius can produce up to a 3 kcal/mol decrease in stability (Figure 4). Both 1.3 Å and 1.2 Å H^{N+} radii (the standard values in AMBER 7[50] and 6[49] GB^{HCT}) overestimate the TIP3P salt bridge stability by as much as 3.8 and 2.4 kcal/mol, respectively. A 3.8 kcal/mol free energy error by itself is on the same order of magnitude as the folding free energies at room temperature[2] and can have profound consequences on the stability of the ion pair and the structural arrangement of residues around it. In comparison, the 1.1 Å profile adequately captures the energy difference between CIP and SSIP, while the intermediate 1.15 Å H^{N+} radii profile comes close to reproducing the CIP→SSIP barrier height, underestimating it by 0.5 kcal/mol. All GB PMFs lack the solvent-separated minima observed with explicit solvent models, resulting in an absence of barrier to salt bridge formation. This limitation of GB, typical of continuum solvent models,[45] stems from the omission of solvent molecularity and is only

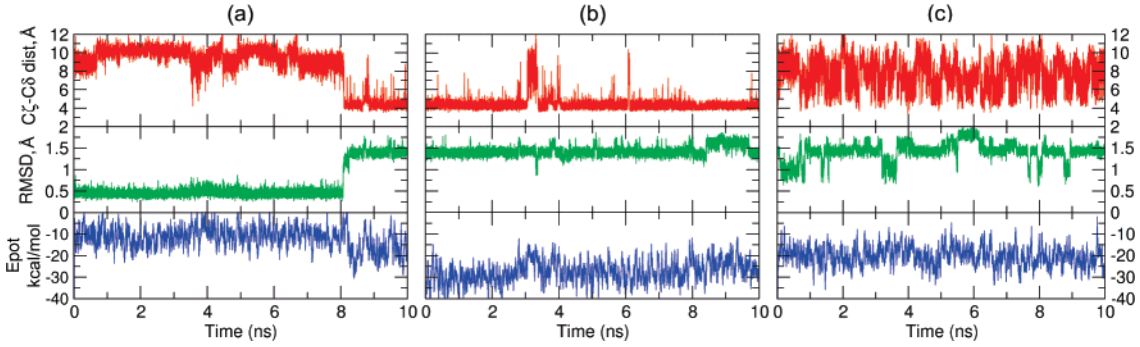


Figure 5. GB^{HCT} simulations of the Fab 17/9 H3 loop with 1.1 Å H^{N+} radii at 300 K, initiated from the X-ray (a) or SB conformations (b). Even with reduced H^{N+} radii, the X-ray loop conformation converts to SB with a 10 kcal/mol decrease in potential energy. The salt bridge, however, reopens transiently in the SB simulation. (c) Simulation from the SB state with neutralized Arg^{97} and Glu^{100} side chains and 1.1 Å H^{N+} radii. Uncharging the salt bridging side chains effectively breaks the salt bridge, but the backbone does not relax back to the native conformation.

addressed in more computationally intensive implicit models such as the RISM formalism,[89−92] molecular surface area-based solvent models,[93,94] or specifically parametrized PB models.[95,96] Employing the recently developed GB^{OBC} model of Onufriev, Bashford, and Case (model II in ref 97) with their suggested Bondi radii[65] modified only for H^N atoms (1.3 Å instead of 1.2 Å) also produced an improved salt bridge profile relative to standard GB^{HCT}, with the PMF underestimating the CIP→SSIP barrier height by 0.5 kcal/mol and overestimating the stabilities of the SSIPs by 0.8 kcal/mol (data not shown, PMF nearly identical to GB^{HCT} H^{N+} = 1.15 Å in Figure 4). As it seems impossible for a simple PDA-based GB model with no depiction of solvent discreteness to capture both the barrier height and the CIP-SSIP energy difference, it seems reasonable to think that GB models should prioritize the correct reproduction of the CIP-SSIP energy gap over the barrier height, since in the absence of solvent discreteness, salt bridge formation is a barrierless downhill process and accurate kinetic behavior cannot be reproduced.

Dynamics were run on the Fab 17/9 H3 loop with H^{N+} radii set to 1.1 Å (Figure 5a), and while the native state could be maintained for an extended period of time (>8ns), the N→SB conversion observed in standard GB^{HCT} still occurs. Because of computational limitations, we could not run a significantly longer or several independent simulations on this system, which would be necessary to fully characterize its kinetic behavior. Yet, a simulation initiated from the SB conformation showed repeated openings of the ion pair, but those events were too transient (<0.5 ns) to allow the loop backbone to relax back to the native conformation (Figure 5c). In contrast, simulations of the SB conformation conducted with the standard radii showed no reopening of the salt bridge (data not shown). Also encouraging was the reduced 10 kcal/mol energy drop accompanying the N→SB transition, down from 14 kcal/mol in standard GB^{HCT} (Figure 1a). This energy difference matches the 4−5 kcal/mol free energy correction visible in the GB^{HCT} 1.1 salt bridge PMF, relative to standard GB^{HCT}, and suggests that additional factors are also responsible for the excessive stability of Fab 17/9 H3 loop non-native conformations. In the following, we take a more systematic approach that is less reliant on backbone parametrization in order to characterize the influence of H^{N+} GB radii on native state stability.

Validation of Radii Modifications on Test Peptides. To assess the relevance of our radii reduction to other systems, including lysine side chains, we studied the PMFs of side-chain ion pairs in small Ala-rich hexapeptides restrained in α-helical conformations. Oppositely charged side chains were spaced one α-helix turn apart (*i*,*i*+4) to create favorable salt-bridge orientations (Figure 6).[98,99] For these simple systems, a stronger positional restraint (10 kcal/mol force constant) was necessary to maintain the backbone in a fully helical conformation.

These exposed salt bridges (Figure 6) displayed markedly reduced stabilities, compared to the Fab 17/9 H3 loop ion pair, due to the absence of a second interacting anionic side chain and the large conformational entropy of the opened state.[100] In particular, the (*i*+4) E,R ion pair, directly comparable to the Fab 17/9 H3 loop ion pair, shows only a 1.2 kcal/mol barrier in TIP3P explicit solvent. This is accompanied by a ∼4 kcal/mol decrease in the CIP-SSIP relative stability from the corresponding pair in Fab 17/9. The same qualitative trend is followed by the GB PMFs, with standard GB^{HCT} still overestimating the stability of the CIP by 2−2.5 kcal/mol, while GB^{HCT} 1.1 falls in close agreement with the TIP3P profile. This improvement suggests that the H^{N+} radii reduction empirically parametrized on the Fab 17/9 salt bridge can be advantageously transferred to other Arg$^+$···Glu$^-$ ion pair geometries.

As observed by Masunov and Lazaridis,[45] the Lys$^+$···Glu$^-$ PMFs tend to be less pronounced, with GB PMFs following, if not accentuating this trend. The discrepancy in interaction energy between GB^{HCT} and TIP3P only fluctuates between 0.7 and 1.6 kcal/mol here, while GB^{HCT} 1.1 falls within 0.5 kcal/mol of the explicit solvent result. This suggests that the GB radii adjustment, while not as crucial as in the stronger Arg$^+$···Glu$^-$ pair, still has the potential to improve the energetics of the Lys$^+$···Glu$^-$ pair appreciably.

Thermodynamical Behavior of the Trp-Cage Miniprotein. As the Fab 17/9 H3 loop native conformation instability appeared to be a coupled salt bridge/backbone problem, we focused our validation effort on Trp-cage TC5b, a miniprotein whose fold has been successfully predicted using long

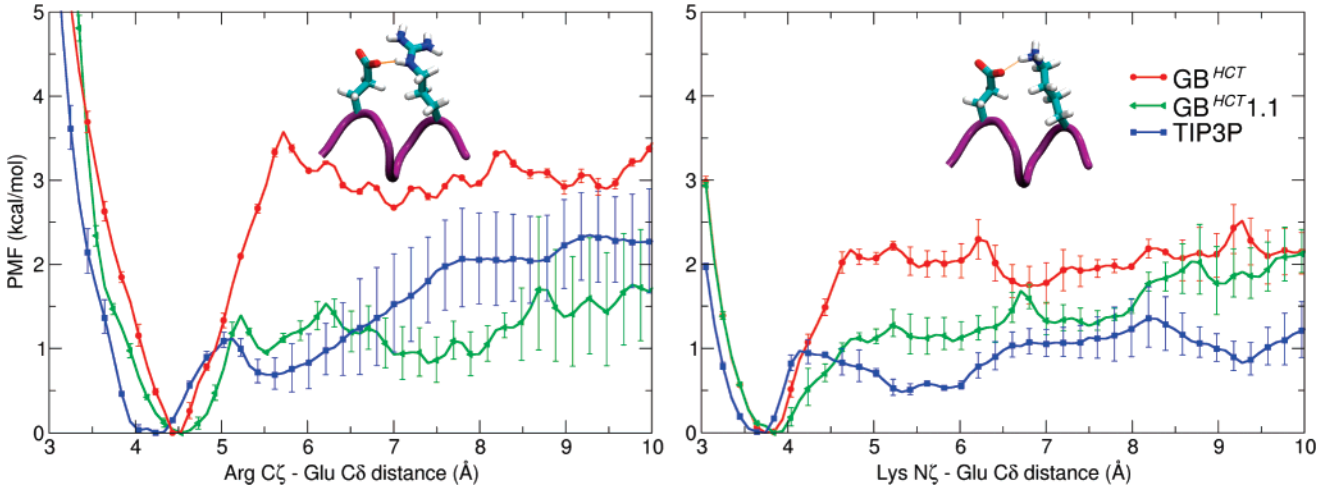


Figure 6. Potentials of mean force of salt bridge formation for the Ac-AEAAARA-NH_2 (left) and Ac-AEAAAKA-NH_2 (right) helical peptides in various solvent models at 300 K. Error bars correspond to separately considering the first or the second half of the data set.

molecular dynamics simulations in implicit solvent,[38] and for which experimental thermodynamic data are available.[101] Even with long MD simulations, there is no assurance that the thermodynamical behavior of protein chains has been sampled to convergence, as some conformational barriers are simply too high to cross on computationally accessible time scales at room temperature. Therefore, we turned to generalized ensemble techniques to evaluate the effect of our GB correction on the thermal stability of charged residue-bearing proteins. However, even generalized ensemble methods such as REMD[44] can require long simulation times to converge, that is why we focused our attention on the TC5b miniprotein construct, a small model (304 atoms), with proteinlike features: a stable fold including tertiary structure and well-defined two state folding kinetics.[101,102] The TC5b construct features an $Arg^{+}\cdots Asp^{-}$ i/i+7 ion pair purposely introduced during the original protein design to generate a stabilizing salt bridge between these positions[101] (Figure 7a). An E5Q mutation was further introduced to avoid forming an unfavorable EXXXD like-charge interaction in the α-helical N-terminal segment of the construct.

Although our lab and others have performed folding simulations of TC5b to near NMR conformations and submitted close to experimental folding rate values,[38,103,104] because of sampling and potential energy accuracy issues, it has proven more challenging to reproduce its full thermodynamic characteristics and in particular experimental melting profiles. The free energy landscape of folding for TC5b has been previously explored by all-atom REMD simulations both in explicit solvent with OPLS-AA[105] by Zhou[106] and implicit solvent using GB^{HCT} and the AMBER ff94[66] force field by Pitera and Swope.[41] Both of these studies predicted significantly higher melting temperatures than the experimental value of 315 K (440 K in TIP3P, ∼400 K in the GB study), raising legitimate doubt about the ability of these force-field/solvation model combinations, parametrized for near-room temperatures, to model temperature-dependent behavior. Additionally, in the implicit solvent study, distorted hydrogen-bonding patterns in solvent-exposed regions of the

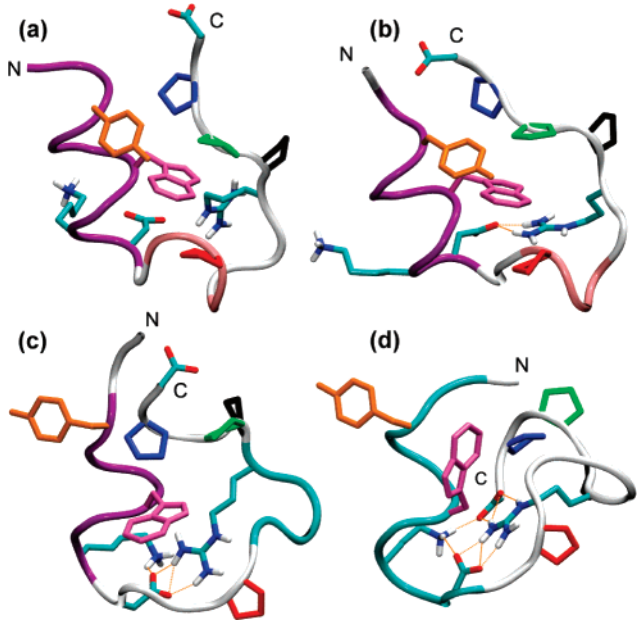


Figure 7. Trp-cage TC5b. (a) Reference NMR structure (model 1 of PDB entry 1L2Y). (b) 267K GB^{HCT} 1.1 REMD global free energy minimum, exhibiting most nativelike features (1.8 Å 3−18 RMSD). (c) 267 K standard GB^{HCT} REMD global minimum and (d) second-lowest free energy minimum, both adopting distorted conformations (2.8 and 3.7 Å 3−18 backbone RMSD, respectively) with multiple salt bridges, not seen in the NMR set. At this temperature in standard GB^{HCT}, the near-NMR ensemble is 1 kcal/mol higher in free energy than the global minimum. The protein backbone is shown in tube representation colored by residue secondary structure type (α: purple, 3_{10}: pink, turn: cyan, coil: white), while Trp-cage motif residue side chains are shown colored as in ref 101: Tyr^{3} (orange), Trp^{6} (magenta), Pro^{12} (red), Pro^{17} (black), Pro^{18} (green), Pro^{19} (blue). Salt bridge forming side chains are colored by atom. H-bonds between ionized side chains are indicated by orange dotted line.

miniprotein were found to cause the largest deviations from the experimental NMR restraints.[41]

Two REMD simulations were performed on the TC5b

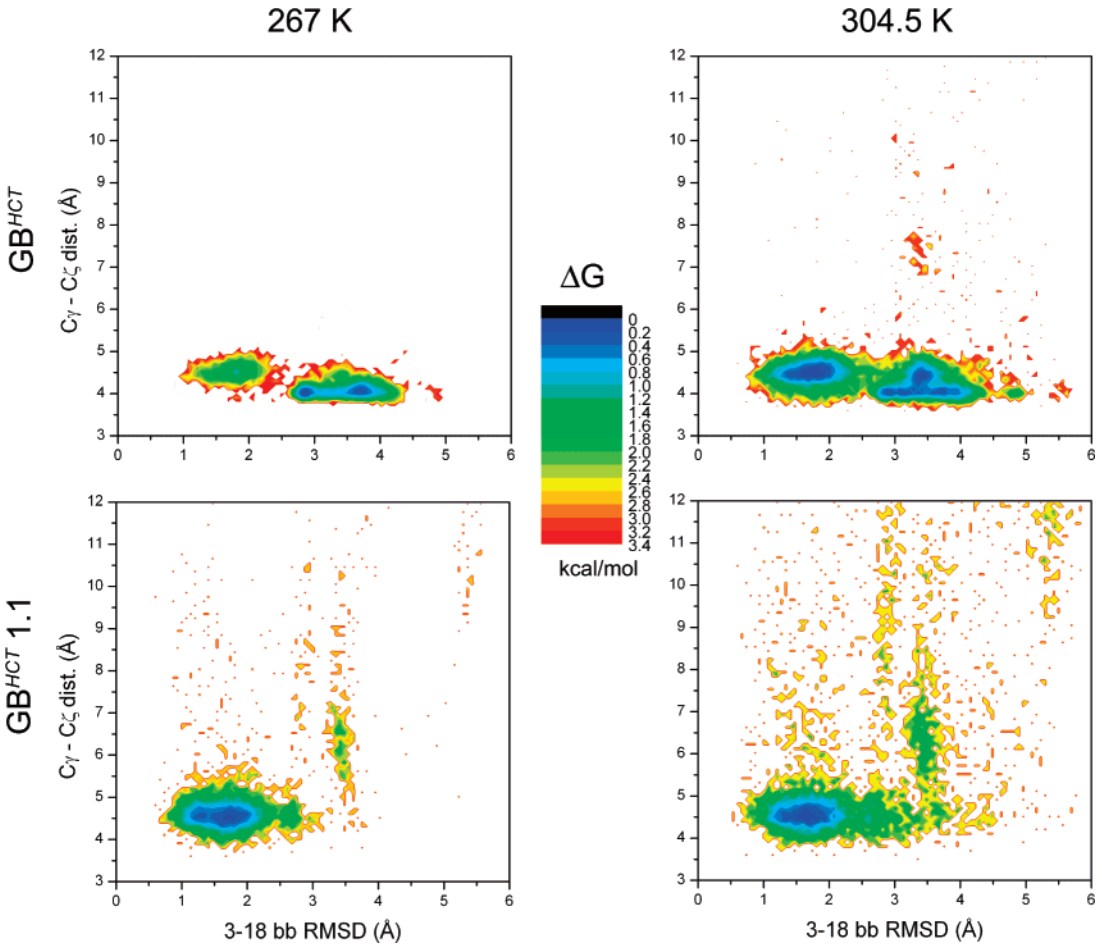


Figure 8. Two-dimensional free energy maps in kcal/mol from REMD data. Top row: GB^{HCT} replicas at 267 K (left) and 304.5 K (right). Bottom row: GB^{HCT} 1.1 replicas at corresponding temperatures. The Asp^{9}···Arg^{16} salt bridge is present in nearly all GB^{HCT} conformations, while it is observed to break more frequently in GB^{HCT} 1.1, particularly for non-native conformations. The average salt bridge distance is *shorter* in the non-native conformations favored by GB^{HCT}. The folded ensemble, which is not the lowest energy basin at 267 K for standard GB^{HCT}, becomes progressively more stable with rising temperature in GB^{HCT}.

construct: one in GB^{HCT} with the standard radii of AMBER8 (modified Bondi[50]), and another with H^{N+} radii set to 1.1 Å, as found optimal for the reproduction of the TIP3P salt bridge PMF in the 17/9 antibody H3 loop study. The simulations covered an extensive temperature range (267−715.7 K) to ensure that high energy barriers did not prevent exhaustive conformational sampling.

Figure 8 shows the TC5b free energy landscapes at various temperatures projected on 2D contour maps using as reaction coordinates the RMSD of backbone heavy atoms in residues 3−18 (corresponding to the well-defined region of the 1L2Y NMR ensemble, with model #1 as reference) and the Asp^{9} Cδ−Arg^{16} Cζ salt bridge distance, to highlight the importance of this ion pair in determining the overall structure of the miniprotein. The lowest free energy basin in standard GB^{HCT} at 267 K (the replica temperature nearest to 0 °C, where experimentally the folded fraction is maximal[101]) comprises only non-native conformations, with a global minimum at 2.8 Å 3−18 RMSD (Figure 7c) and an almost equiprobable ($\Delta G \sim 0.13$ kcal/mol) other minimum at 3.7 Å (Figure 7d). Cluster analysis on the structures comprising this unfolded basin reveals dominant, enthalpically favored ionic networks involving nearly all formally charged moieties of the miniprotein (C-terminal carboxylate, Lys^{8}, Asp^{9}) clustering around Arg^{16} (Figure 7c,d). Once again, these formations underline the insufficient desolvation penalty incurred by charged groups in the standard GB^{HCT} model, as the NMR ensemble shows only one such ionic interaction: the Asp^{9}···Arg^{16} salt bridge. In addition, the Arg^{16} side chain is not well resolved by the NMR assignment,[38] suggesting ample conformational freedom, incompatible with the rigid and thermodynamically stable ionic networks observed in Figure 7c,d.

At 267 K, while the folded region (RMSD < 2.5 Å) is 1 kcal/mol higher in free energy than the global minimum in standard GB^{HCT}, it is the lowest free energy basin in GB^{HCT} 1.1, with a free energy minimum at 1.8 Å RMSD and 4.5 Å salt bridge distance. Structures in this basin still display most of the features of the TC5b native fold: a 2−8 helix (mainly α-helical, according to DSSP[107]) and a second helical segment between residues 11 and 14, with equal proportions of 3_{10}- and α-helical conformations. As a representative structure of the GB^{HCT} 1.1 267 K free energy minimum ensemble shows (Figure 7b), the largest deviation from the NMR reference conformation (Figure 7a) occurs between the 3_{10}-helix and the C-terminal polyproline II segment at Ser^{14} and the flexible Gly^{15}, inducing a slight shift in the location of the polyproline II helix. Nevertheless, the key

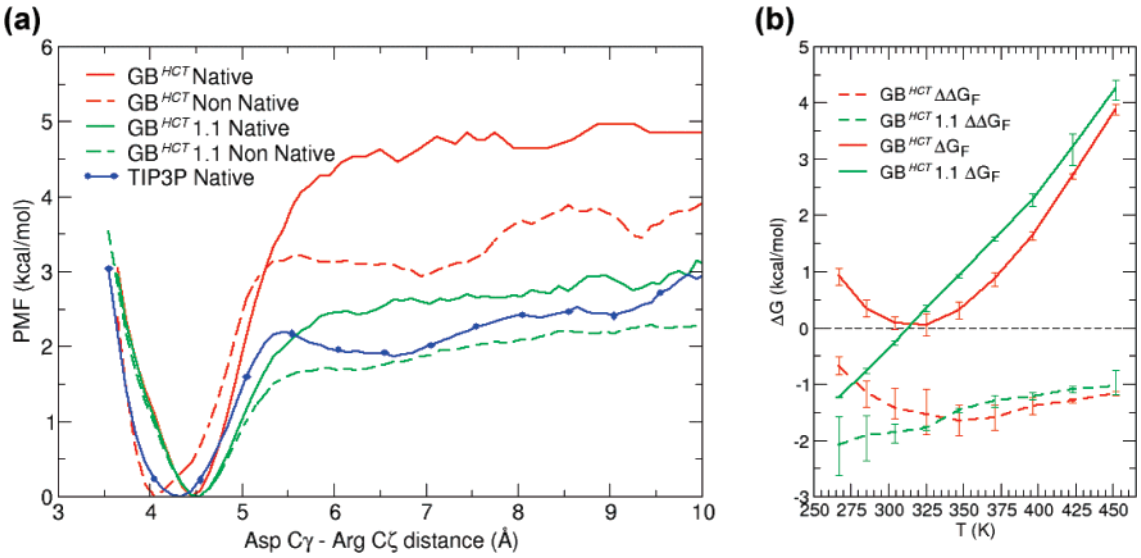


Figure 9. (a) Salt bridge formation PMFs along the Asp[9] Cγ−Arg[16] Cζ distance coordinate for native and non-native ensembles at 304.5 K from REMDs in GB with different radii sets or TIP3P water at 306 K. All curves were smoothed by taking 4-point moving window averages. (b) Native state stabilization by salt bridge, $\Delta\Delta G_F$, as calculated from eq 5 (dashed lines), and folding free energy, ΔG_F (solid lines), in GB^{HCT} with 1.3 Å (red) and 1.1 Å (green) H^{N+} radii. Error bars correspond to separately considering the first or the second half of the data set.

hydrophobic cage motif remains well preserved, with Tyr[3], Trp[6], Pro[12], Pro[18], and Pro[19] clustering nearly as well as in the NMR models. As in the NMR models, the Asp[9]···Arg[16] salt bridge is present, while the Lys[8] side chain is fully solvent exposed and does not take part in intramolecular ionic interactions.

Interestingly, in standard GB^{HCT}, the salt bridge distance distribution is shifted toward longer distances by almost 0.5 Å in the low RMSD ensemble, reflecting the preferential formation of a monodentate H-bond in the near-NMR ensemble ion pair (Figure 7b), while in distorted low free energy structures, a bidentate interaction is favored by the recruitment of Lys[8] in the ionic network (Figure 7c,d). The relative stability of these distorted low-energy structures quickly decreases with temperature, since already at 304.5 K, the properly folded ensemble is more stable by 0.08 kcal/mol. However, the most striking difference between the two GB parameter sets lies in the composition of their respective non-native ensembles.

Salt bridge formation PMFs were calculated at 304.5 K from the GB1.3 and GB1.1 REMD populations and compared to one derived from a short (26 ns) REMD in TIP3P explicit solvent started from the NMR-derived conformation (Figure 9a). This simulation time is sufficient to effectively sample salt bridge distances in the folded state. More details on this explicit solvent REMD simulation will be published elsewhere. The GB^{HCT} 1.1 salt bridge PMF for the native ensemble shows much improved agreement with its TIP3P counterpart, as it marginally overestimates the SSIP stability by ∼0.7 kcal/mol, while standard GB^{HCT} overestimates it by as much as ∼2.7 kcal/mol.

To follow the influence of ion pair strength on the overall stability of the TC5b miniprotein, we calculated the stability contribution of the salt bridge to the folding free energy, $\Delta\Delta G_F$, defined as the difference in free energy of folding of the protein with and without the salt bridge.[108] This quantity corresponds to the free energy difference between forming the salt bridge in the folded and unfolded states:

$$\Delta\Delta G_F = \Delta G_F^{sb} - \Delta G_F^{nosb} = (G_F^{sb} - G_U^{sb}) - (G_F^{nosb} - G_U^{nosb}) = (G_U^{nosb} - G_U^{sb}) - (G_F^{nosb} - G_F^{sb}) \quad (5)$$

Therefore, assuming REMD generates a converged thermodynamic ensemble at each of the replica temperatures, $\Delta\Delta G_F$ was calculated for all REMD temperatures by

$$\Delta\Delta G_F = -RT\ln\left(\frac{p_U^{nosb}\cdot p_F^{sb}}{p_U^{sb}\cdot p_F^{nosb}}\right) \quad (6)$$

where p_U and p_F stand for the unfolded and folded fractions in the absence or presence of the Asp[9]···Arg[16] salt bridge (nosb and sb superscripts, respectively). An RMSD cutoff of 2.5 Å was adopted to define the folded state as it clearly corresponds to the peak of the barrier between native and non-native basins in the 2D PMFs (Figure 8). We defined salt bridged states as having an Asp[9] Cγ−Arg[16] Cζ distance ≤5.5 Å. This value corresponds approximately to the peak of the CIP→SSIP transition barrier in the explicit solvent salt bridge PMF (Figure 9a).

Figure 9b simultaneously shows $\Delta\Delta G_F$ and the overall folding free energy, ΔG_F, for both the standard GB^{HCT} and GB^{HCT} 1.1 REMD simulations. At 304.5 K, the salt bridge stabilization of the folded state is ∼1.4 kcal/mol in GB^{HCT} and ∼1.8 kcal/mol in GB^{HCT} 1.1, well within the range of experimentally observed values.[109]

At low temperatures, the Asp[9]···Arg[16] salt bridge in standard GB^{HCT}, although intrinsically stronger (Figure 9a), actually stabilizes the native state *less* than in GB^{HCT} 1.1. This phenomenon stems from the almost equally strong stabilization of non-native conformations by the salt bridge, as observed in the two compact 267 K GB^{HCT} free energy minima (Figure 7c,d). On the contrary, past 340 K, the salt

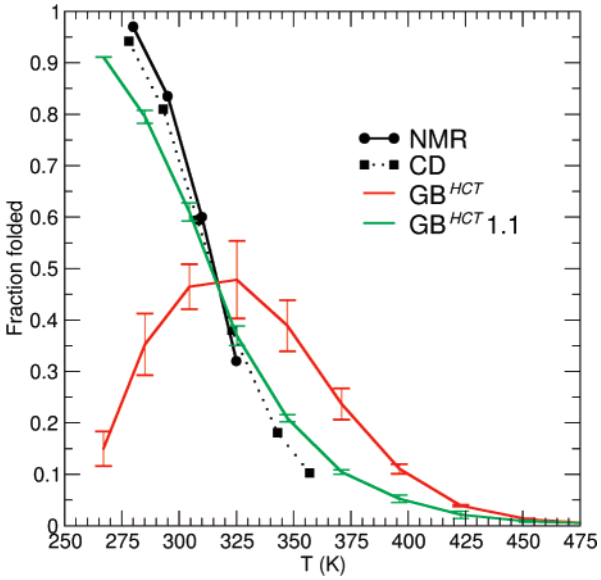


Figure 10. Experimental (CD, NMR CSD)[101] and simulated melting curves for the TC5b miniprotein. Error bars estimating the sampling uncertainty were determined by considering separately the first and second halves of the data set, each 45 ns in GB^{HCT}, and 25 ns in GB^{HCT} 1.1.

bridge contributes more strongly to folded state stability in standard GB^{HCT} than in GB^{HCT} 1.1, at least partly accounting for the increased high-T stability of TC5b in standard GB^{HCT}.

Interestingly, in both GB^{HCT} and the improved GB^{HCT} 1.1 solvent models, salt bridge stabilization decreases much more slowly than the overall protein stability with increasing temperature and remains stabilizing at elevated temperatures, providing a rationale for the increased number of ionic pairs observed in proteins from hyperthermophilic organisms.[1,10,12−15] In addition, the GB^{HCT} model used here, with its constant water dielectric of 78.5, is expected to provide only an underestimation of salt bridge stability at high temperature, as experimentally the dielectric constant of water continuously decreases with temperature, to reach only ∼55 at 100 °C and 1 atm,[110] thus favoring Coulombic interactions even more at high temperatures.[111]

Finally, still using a folding criterion of 3−18 backbone RMSD ≤ 2.5 Å, it is possible to generate melting curves for TC5b in each of the GB models and compare them to experiment (Figure 10). The standard GB^{HCT} produces a melting profile shifted upward by ∼30 K relative to the NMR and CD experimental profiles, as reported by Pitera and Swope.[41] In addition, the preponderance of enthalpically stabilized non-native structures is responsible for the drop in folded fraction at low-temperature seen in the GB^{HCT} profile and even prevents reaching the melting transition midpoint (Figure 10). In sharp contrast, the melting temperature measured by cubic spline interpolation of the melting profile in GB^{HCT} 1.1 (314 K) is in excellent agreement with the experimental value of 315 K.[101] Furthermore, the entire simulated melting profile falls in very good overall agreement with experiment, only departing noticeably (∼10% of fraction folded) from the experimental curves at extremes of the temperature range. We therefore concur with the earlier conclusions by Zhou[106] and Pitera[41] that current force fields are most accurate around room temperature, where they were parametrized. This is especially true of their GB component, which most commonly fails to include the temperature dependence of the dielectric constant. However, current force fields ought to be able to at least predict near-room-temperature melting temperatures, and we show here that in GB^{HCT} this ability was only obscured by the overwhelming influence of incorrectly treated ionic interactions.

Since we did not have to parametrize our potential function against variable temperature data to capture the most important features of the melting profile and a correct T_m value, we believe that the large overapproximations of T_m by earlier REMD studies[41,106] might originate from insufficient sampling caused by overwhelming salt bridges in GB—as was the case in our GB^{HCT} run with standard H^{N+} radii—or solvent friction in TIP3P, preventing reaching ergodicity in either case. In the Pitera and Swope study, the mostly helical TC5b native structure might also have drawn stability from the helical bias of the AMBER ff94 force field[112] that they employed. Although Pitera and Swope employed an energy function very similar to ours (AMBER ff94/GB^{HCT}), they did not report observing the distorted structures with ionic networks we encountered in GB^{HCT}. Possible reasons for this discrepancy could be the small number of replica exchanges attempted in their protocol (400), hampering the sampling of remote regions of the folding landscape by low-temperature replicas, or the excessive stability of helices in ff94 preventing formation of non-native conformations.

Conclusions

Simple GB models based on the pairwise descreening approximation are a popular choice for molecular simulations as they allow significant computational speedup over more accurate GB implementations or PB equation-based implicit solvent models.[113] It is therefore important to ensure that they can achieve an adequate level of accuracy.

Using potentials of mean force, we were able to quantify the problematic overstabilization of ionic pairs observed in the standard GB^{HCT} implementation of the AMBER package. A simple empirical reduction of the GB radii of H^{N+} atoms from 1.3 to 1.1 Å allows a close reproduction of explicit solvent CIP-SSIP relative energies in both the Fab 17/9 H3 loop Arg^{97}···Glu^{100} ion pair and test model helical peptide systems. We note that this ad hoc modification of a single intrinsic Born radius should be followed by a more complete assessment of the influence of all of the radii on the system thermodynamics. In the absence of solvent discreteness, salt bridge formation also remains a kinetically downhill process, and therefore GB models cannot be expected to accurately reproduce the kinetics of conformational transitions. This shortcoming has started to be addressed,[21,22] notably by explicitly including the first solvation shell around solutes in mixed explicit/implicit solvent models.[114−120]

By comparing experimental thermal denaturation data to REMD simulations of the Trp-cage miniprotein TC5b, we confirmed that charge−charge interactions clearly outweigh the desolvation penalty incurred by ionized side chains upon salt bridge formation in the standard GB^{HCT} model of

AMBER. In sharp contrast, the same GB model with only reduced H^{N+} Born radii closely captures the thermodynamics of the $Asp^{9}\cdots Arg^{16}$ salt bridge and for the first time allowed the generation of a near-experimental melting profile for TC5b. The GB^{HCT} 1.1 T_m value of 314 K falls in remarkable agreement with the experimental value of 315 K, while the standard GB^{HCT} profile approaches the melting transition midpoint but shows a disturbing preference for non-native structures at low temperature. While the GB model should at least be capable of correctly predicting thermodynamic observables at or around room temperature, we believe that our accurate reproduction of the TC5b melting profile likely arises from fortuitous cancellation of error, as GB^{HCT} was not parametrized to reproduce the temperature dependence of water solvation.

This study also provides further indication that strong electrostatic interactions are not predominant factors in protein native state stability, as they can stabilize non-native states by similar or even greater amounts, depending on the unfolded state topology.[9,109] Nevertheless, from our modified GB REMD simulation of the TC5b miniprotein, it appears that native state stabilization by ionic interactions decreases at a slower rate than the overall protein stability with increasing temperature, providing a rationale for the observed preponderance of charged amino acid residues in the proteins of thermophilic organisms.[1,10,12−15]

Acknowledgment. Supercomputer time on the NCSA Platinum and Tungsten Linux Clusters (NCSA MCA02N028) and financial support from the National Institutes of Health (NIH GM6167803) are gratefully acknowledged. Additional computer time on a large Altix was generously provided by the SGI Engineering group. C.S. is a Cottrell Scholar of Research Corporation. R. Geney is grateful to the Chemical Computing Group and ACS COMP division for a CCG Excellence Award. The authors also thank Alan Grossfield for making his WHAM code available and David Case, Alexey Onufriev, and John Mongan for helpful discussions.

Supporting Information Available: Full citation for ref 52. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Perutz, M. F. *Science* **1978**, *201*, 1187−1191.

(2) Dill, K. A. *Biochemistry* **1990**, *29*, 7133−7155.

(3) Honig, B.; Yang, A. S. *Adv. Protein Chem.* **1995**, *46*, 27−58.

(4) Honig, B.; Hubbell, W. L. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 5412−5416.

(5) Daopin, S.; Soderlind, E.; Baase, W. A.; Wozniak, J. A.; Sauer, U.; Matthews, B. W. *J. Mol. Biol.* **1991**, *221*, 873−887.

(6) Hendsch, Z. S.; Tidor, B. *Protein Sci.* **1994**, *3*, 211−226.

(7) Waldburger, C. D.; Schildbach, J. F.; Sauer, R. T. *Nat. Struct. Biol.* **1995**, *2*, 122−128.

(8) Waldburger, C. D.; Jonsson, T.; Sauer, R. T. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 2629−2634.

(9) Dong, F.; Zhou, H. X. *Biophys. J.* **2002**, *83*, 1341−1347.

(10) Elcock, A. H. *J. Mol. Biol.* **1998**, *284*, 489−502.

(11) Xiao, L.; Honig, B. *J. Mol. Biol.* **1999**, *289*, 1435−1444.

(12) Vieille, C.; Zeikus, G. J. *Microbiol. Mol. Biol. Rev.* **2001**, *65*, 1−43.

(13) Zhou, H. X. *Biophys. J.* **2002**, *83*, 3126−3133.

(14) Dominy, B. N.; Minoux, H.; Brooks, C. L. *Proteins* **2004**, *57*, 128−141.

(15) Thomas, A. S.; Elcock, A. H. *J. Am. Chem. Soc.* **2004**, *126*, 2208−2214.

(16) Karplus, M.; McCammon, J. A. *Nat. Struct. Biol.* **2002**, *9*, 646−652.

(17) Cramer, C. J.; Truhlar, D. G. *Chem. Rev.* **1999**, *99*, 2161−2200.

(18) Roux, B.; Simonson, T. *Biophys. Chem.* **1999**, *78*, 1−20.

(19) Simonson, T. *Curr. Opin. Struct. Biol.* **2001**, *11*, 243−252.

(20) Simonson, T. *Rep. Prog. Phys.* **2003**, *66*, 737−787.

(21) Feig, M.; Brooks, I., Charles L *Curr. Opin. Struct. Biol.* **2004**, *14*, 217−224.

(22) Baker, N. A. *Curr. Opin. Struct. Biol.* **2005**, *15*, 137−143.

(23) Kramers, H. A. *Physica* **1940**, *7*, 284−304.

(24) Weiner, J. H.; Pear, M. R. *Macromolecules* **1977**, *10*, 317−325.

(25) Chandler, D. *J. Chem. Phys.* **1978**, *68*, 2959−2970.

(26) Helfand, E. *Physica A* **1983**, *118*, 123−135.

(27) Ansari, A.; Jones, C. M.; Henry, E. R.; Hofrichter, J.; Eaton, W. A. *Science* **1992**, *256*, 1796−1798.

(28) Haran, G.; Haas, E.; Rapaport, D. C. *J. Phys. Chem.* **1994**, *98*, 10294−10302.

(29) Takano, M.; Yamato, T.; Higo, J.; Suyama, A.; Nagayama, K. *J. Am. Chem. Soc.* **1999**, *121*, 605−612.

(30) Karplus, M. *J. Phys. Chem. B* **2000**, *104*, 11−27.

(31) Zagrovic, B.; Pande, V. *J. Comput. Chem.* **2003**, *24*, 1432−1436.

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

(33) Guillot, B. *J. Mol. Liq.* **2002**, *101*, 219−260.

(34) Constanciel, R.; Contreras, R. *Theor. Chim. Acta* **1984**, *65*, 1−11.

(35) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127−6129.

(36) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. *J. Phys. Chem. A* **1997**, *101*, 3005−3014.

(37) Cornell, W.; Abseher, R.; Nilges, M.; Case, D. A. *J. Mol. Graph.* **2001**, *19*, 136−145.

(38) Simmerling, C.; Strockbine, B.; Roitberg, A. E. *J. Am. Chem. Soc.* **2002**, *124*, 11258−11259.

(39) Zhou, R. H. *Proteins* **2003**, *53*, 148−161.

(40) Zhou, R. H.; Berne, B. J. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12777−12782.

(41) Pitera, J. W.; Swope, W. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 7587−7592.

(42) Felts, A. K.; Harano, Y.; Gallicchio, E.; Levy, R. M. *Proteins* **2004**, *56*, 310−321.

(43) Ghosh, A.; Rapp, C. S.; Friesner, R. A. *J. Phys. Chem. B* **1998**, *102*, 10983−10990.

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

(45) Masunov, A.; Lazaridis, T. *J. Am. Chem. Soc.* **2003**, *125*, 1722−1730.

(46) Dominy, B. N.; Brooks, C. L. *J. Phys. Chem. B* **1999**, *103*, 3765−3773.

(47) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *Chem. Phys. Lett.* **1995**, *246*, 122−129.

(48) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1996**, *100*, 19824−19839.

(49) Tsui, V.; Case, D. A. *J. Am. Chem. Soc.* **2000**, *122*, 2489−2498.

(50) Tsui, V.; Case, D. A. *Biopolymers* **2001**, *56*, 275−291.

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

(52) Case, D. A. et al. AMBER 8, University of California, San Francisco, 2004.

(53) Wang, J. M.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049−1074.

(54) Beachy, M. D.; Chasman, D.; Murphy, R. B.; Halgren, T. A.; Friesner, R. A. *J. Am. Chem. Soc.* **1997**, *119*, 5908−5920.

(55) Hornak, V.; Roitberg, A. E.; Simmerling, C. Manuscript in preparation.

(56) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327−341.

(57) Essex, J. W. *Mol. Simul.* **1998**, *20*, 159−178.

(58) Hochtl, P.; Boresch, S.; Bitomsky, W.; Steinhauser, O. *J. Chem. Phys.* **1998**, *109*, 4927−4937.

(59) Richardi, J.; Fries, P. H.; Millot, C. *J. Mol. Liq.* **2005**, *117*, 3−16.

(60) Rini, J. M.; Schulze-Gahmen, U.; Wilson, I. A. *Science* **1992**, *255*, 959−965.

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

(62) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577−8593.

(63) Crowley, M. F.; Darden, T. A.; Cheatham, T. E.; Deerfield, D. W. *J. Supercomput.* **1997**, *11*, 255−278.

(64) Toukmaji, A.; Sagui, C.; Board, J.; Darden, T. *J. Chem. Phys.* **2000**, *113*, 10913−10927.

(65) Bondi, A. *J. Phys. Chem.* **1964**, *68*, 441−&.

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

(67) Torrie, G. M.; Valleau, J. P. *J. Comput. Phys.* **1977**, *23*, 187−199.

(68) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011−1021.

(69) Roux, B. *Comput. Phys. Commun.* **1995**, *91*, 275−282.

(70) Hansmann, U. H. E. *Chem. Phys. Lett.* **1997**, *281*, 140−150.

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

(72) Hornak, V.; Simmerling, C. *J. Mol. Graph.* **2004**, *22*, 405−413.

(73) Hornak, V.; Simmerling, C. *Proteins* **2003**, *51*, 577−590.

(74) Kabat, E. A.; Wu, T. T.; Perry, H. M.; Gottesman, K. S.; C., F. *Sequences of Proteins of Immunological Interest*; Diane Books Publishing Company: 1991.

(75) Bader, J. S.; Chandler, D. *J. Phys. Chem.* **1992**, *96*, 6423−6427.

(76) Rey, R.; Guardia, E. *J. Phys. Chem.* **1992**, *96*, 4712−4718.

(77) Friedman, R. A.; Mezei, M. *J. Chem. Phys.* **1995**, *102*, 419−426.

(78) Resat, H.; Mezei, M.; McCammon, J. A. *J. Phys. Chem.* **1996**, *100*, 1426−1433.

(79) Martorana, V.; La Fata, L.; Bulone, D.; San Biagio, P. L. *Chem. Phys. Lett.* **2000**, *329*, 221−227.

(80) Rozanska, X.; Chipot, C. *J. Chem. Phys.* **2000**, *112*, 9691−9694.

(81) Belch, A. C.; Berkowitz, M.; McCammon, J. A. *J. Am. Chem. Soc.* **1986**, *108*, 1755−1761.

(82) Gruia, A. D.; Fischer, S.; Smith, J. C. *Proteins* **2003**, *50*, 507−515.

(83) Gruia, A. D.; Fischer, S.; Smith, J. C. *Chem. Phys. Lett.* **2004**, *385*, 337−340.

(84) Born, M. *Z. Phys.* **1920**, *1*, 45−48.

(85) Dudek, M. J.; Ponder, J. W. *J. Comput. Chem.* **1995**, *16*, 791−816.

(86) Barlow, D. J.; Thornton, J. M. *J. Mol. Biol.* **1983**, *168*, 867−885.

(87) Kumar, S.; Nussinov, R. *J. Mol. Biol.* **1999**, *293*, 1241−1255.

(88) Zhu, J.; Alexov, E.; Honig, B. *J. Phys. Chem. B* **2005**, *109*, 3008−3022.

(89) Chandler, D. *Annu. Rev. Phys. Chem.* **1978**, *29*, 441−471.

(90) Hirata, F.; Rossky, P. J. *Chem. Phys. Lett.* **1981**, *83*, 329−334.

(91) Hirata, F.; Rossky, P. J.; Pettitt, B. M. *J. Chem. Phys.* **1983**, *78*, 4133−4144.

(92) Pettitt, B. M.; Rossky, P. J. *J. Chem. Phys.* **1986**, *84*, 5836−5844.

(93) Fukunishi, Y.; Suzuki, M. *J. Phys. Chem.* **1996**, *100*, 5634−5636.

(94) Fukunishi, Y.; Suzuki, M. *J. Comput. Chem.* **1997**, *18*, 1656−1663.

(95) Rashin, A. A. *J. Phys. Chem.* **1989**, *93*, 4664−4669.

(96) Pratt, L. R.; Hummer, G.; Garcia, A. E. *Biophys. Chem.* **1994**, *51*, 147−165.

(97) Onufriev, A.; Bashford, D.; Case, D. A. *Proteins* **2004**, *55*, 383−394.

(98) Marqusee, S.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 8898−8902.

(99) Perutz, M. F.; Fermi, G. *Proteins* **1988**, *4*, 294−295.

(100) Luo, R.; David, L.; Hung, H.; Devaney, J.; Gilson, M. K. *J. Phys. Chem. B* **1999**, *103*, 727−736.

(101) Neidigh, J. W.; Fesinmeyer, R. M.; Andersen, N. H. *Nat. Struct. Biol.* **2002**, *9*, 425−430.

(102) Qiu, L. L.; Pabit, S. A.; Roitberg, A. E.; Hagen, S. J. *J. Am. Chem. Soc.* **2002**, *124*, 12952−12953.

(103) Snow, C. D.; Zagrovic, B.; Pande, V. S. *J. Am. Chem. Soc.* **2002**, *124*, 14548−14549.

(104) Chowdhury, S.; Lee, M. C.; Xiong, G. M.; Duan, Y. *J. Mol. Biol.* **2003**, *327*, 711−717.

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

(106) Zhou, R. H. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13280−13285.

(107) Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577−2637.

(108) Bosshard, H. R.; Marti, D. N.; Jelesarov, I. *J. Mol. Recognit.* **2004**, *17*, 1−16.

(109) Pace, C. N.; Alston, R. W.; Shaw, K. L. *Protein Sci.* **2000**, *9*, 1395−1398.

(110) Fernandez, D. P.; Goodwin, A. R. H.; Lemmon, E. W.; Sengers, J.; Williams, R. C. *J. Phys. Chem. Ref. Data* **1997**, *26*, 1125−1166.

(111) Elcock, A. H.; McCammon, J. A. *J. Phys. Chem. B* **1997**, *101*, 9624−9634.

(112) Cornell, W. D.; Caldwell, J. W.; Kollman, P. A. *J. Chim. Phys.-Chim. Biol.* **1997**, *94*, 1417−1435.

(113) Feig, M.; Onufriev, A.; Lee, M. S.; Im, W.; Case, D. A.; Brooks, C. L. *J. Comput. Chem.* **2004**, *25*, 265−284.

(114) Beglov, D.; Roux, B. *J. Chem. Phys.* **1994**, *100*, 9050−9063.

(115) Lounnas, V.; Ludemann, S. K.; Wade, R. C. *Biophys. Chem.* **1999**, *78*, 157−182.

(116) Topol, I. A.; Tawa, G. J.; Burt, S. K.; Rashin, A. A. *J. Chem. Phys.* **1999**, *111*, 10998−11014.

(117) Rosenhouse-Dantsker, A.; Osman, R. *Biophys. J.* **2000**, *79*, 66−79.

(118) Lee, M. S.; Salsbury, F. R.; Olson, M. A. *J. Comput. Chem.* **2004**, *25*, 1967−1978.

(119) Kentsis, A.; Mezei, M.; Osman, R. *Biophys. J.* **2003**, *84*, 805−815.

(120) Yu, Z. Y.; Jacobson, M. P.; Josovitz, J.; Rapp, C. S.; Friesner, R. A. *J. Phys. Chem. B* **2004**, *108*, 6643−6654.

CT050183L

JCTC Journal of Chemical Theory and Computation

Estimation of Absolute Free Energies of Hydration Using Continuum Methods: Accuracy of Partial Charge Models and Optimization of Nonpolar Contributions

Robert C. Rizzo,*[,†,§] Tiba Aynechi,[†,||] David A. Case,[‡] and Irwin D. Kuntz[†]

Department of Pharmaceutical Chemistry, University of California at San Francisco, San Francisco, California 94143-2240, and the Department of Molecular Biology TPC-15, The Scripps Research Institute, La Jolla, California 92037

Received April 12, 2005

Abstract: Absolute free energies of hydration (ΔG_{hyd}) for more than 500 neutral and charged compounds have been computed, using Poisson−Boltzmann (PB) and Generalized Born (GB) continuum methods plus a solvent-accessible surface area (SA) term, to evaluate the accuracy of eight simple point-charge models used in molecular modeling. The goal is to develop improved procedures and protocols for protein−ligand binding calculations and virtual screening (docking). The best overall PBSA and GBSA results, in comparison with experimental ΔG_{hyd} values for small molecules, were obtained using MSK, RESP, or ChelpG charges obtained from ab initio calculations using 6-31G* wave functions. Correlations using semiempirical (AM1BCC, AM1CM2, and PM3CM2) or empirical (Gasteiger-Marsili and MMFF94) methods yielded mixed results, particularly for charged compounds. For neutral compounds, the AM1BCC method yielded the best agreement with experimental results. In all cases, the PBSA and GBSA results are highly correlated (overall $r^2 = 0.94$), which highlights the fact that various partial charge models influence the final results much more than which continuum method is used to compute hydration free energies. Overall improved agreement with experimental results was demonstrated using atom-based constants in place of a single surface area term. Sets of optimized SA constants, suitable for use with a given charge model, were derived by fitting to the difference in experimental free energies and polar continuum results. The use of optimized atom-based SA constants for the computation of ΔG_{hyd} can fine-tune already reasonable agreement with experimental results, ameliorate gross deficiencies in any particular charge model, account for nonoptimal radii, or correct for systematic errors.

Introduction

The quantification of how a solute will partition into two different phases, A and B, is widely used in drug design.[1,2] Notable examples include using *n*-octanol-to-water partitioning ($\log P_{octanol/water}$) as a model for cell membrane permeability and gas-to-water partitioning ($\log P_{gas/water}$) to estimate desolvation penalties associated with protein−ligand binding. The two quantities are related, from the perspective of continuum models of solvation, in that they quantify partitioning between phases with low (gas ∼ 1 and octanol ∼ 17) and high (water ∼ 80) dielectric constants. Experimental $\log P_{gas/water}$ measurements, often expressed as free energies of hydration ($\Delta G_{hyd} = -2.3RT \log P_{gas/water}$), have been compiled by several research groups for both neutral and charged species (see Table S1 in the Supporting

* Corresponding author e-mail: rizzo@ams.sunysb.edu; phone: (415) 902-3549.
† University of California at San Francisco.
‡ The Scripps Research Institute.
§ Current address: Department of Applied Mathematics and Statistics, Stony Brook University, Stony Brook, NY 11794-3600.
|| Graduate Group in Biophysics.

10.1021/ct050097l CCC: $33.50
Published on Web 11/15/2005

Information).[3−9] These experimental data make the computation of ΔG_{hyd} an attractive thermodynamic property for validating continuum simulation methods and can be used to guide the choice of parameters employed in such calculations.

The ultimate goal of this study is to optimize computational methods for protein−ligand binding calculations and virtual screening (docking). The recently reported Molecular Mechanics Poisson−Boltzmann Surface Area (MM-PBSA) and Molecular Mechanics Generalized Born Surface Area (MM-GBSA) methods[10−12] incorporate a ΔG_{hyd}-like term as a measure of the change in desolvation ($\Delta\Delta G_{hyd}$) for the receptor−ligand binding event.[13−20] In MM-PBSA and MM-GBSA analysis,[10−12] PBSA or GBSA continuum energy terms for a given species (complex, receptor, or ligand) are formally equivalent to an absolute ΔG_{hyd} if, as is commonly done, dielectric constants of 1 (gas phase) and 80 (water phase) are specified. Therefore, the accuracy of computed ΔG_{hyd} terms directly affect the final computed binding energies. Unfortunately, experimental free energies of hydration are not available for proteins, most drugs, or protein−drug complexes. A reasonable alternative is to verify that the calculation methods and parameters yield good results for small organic molecules, for which experimental absolute free energies of hydration are available,[3−9] prior to using MM-PBSA and MM-GBSA methods.

Historically, the most accurate ΔG_{hyd} calculations have employed free energy perturbation (FEP) or thermodynamic integration (TI) simulations incorporating explicit models of water.[21,22] This was first done in 1985 by Jorgensen and Ravimohan[23] who used FEP methods to compute the relative free energy of hydration ($\Delta\Delta G_{hyd}$) for ethane and methanol in excellent agreement with experimental results using Monte Carlo simulations. The FEP and TI methods yield $\Delta\Delta G_{hyd}$ (or ΔG_{hyd}) directly and without the need for partitioning the free energy into separate components, as in other more-approximate approaches. However, such simulations can be tedious to set up and too computationally expensive for high-throughput structure-based drug design.

Continuum theories which treat solvent as a bulk macroscopic quantity[24] represent a complementary approach to the computation of solute hydration. In particular, Poisson−Boltzmann (PB)[25] and Generalized Born (GB)[26] are two widely used methods used to estimate the polarization energy associated with bringing any species from the gas phase to the bulk solvent phase. PB and GB calculation results are typically augmented by a solvent-accessible surface area term (SA) to account for nonpolar contributions to the total free energy of hydration. A comprehensive study which compares the performance of various GB implementations to PB reference calculations has recently been reported by Feig et al.[27] In this paper, we instead focus on evaluating which commonly used partial charge models yield GBSA and PBSA absolute hydration free energies in agreement with experimental results.

Two early continuum studies that directly compare computed ΔG_{hyd} with experimental results include the original GBSA report by Still et al.[26] and the Sitkoff et al.[25] PARSE (parameters for solvation energy) study designed for use with PBSA methods. Excellent results were obtained in both cases; however, the number of molecules tested was relatively small (between 20 and 67 molecules).[25,26] Both prior studies employed charge models based on functional group assignment, which may be difficult to assign to compounds typically found in databases used for high-throughput virtual screening (docking). More recent efforts have focused on evaluating the accuracy of partial charge models that may be more easily assigned, in an automated fashion, to relatively large and diverse data sets.[4,28−31] For example, Jorgensen and co-workers have recently reported the implementation and validation of a generalized GBSA model in conjunction with the OPLS-AA force field employing charges obtained from AM1CM1 semiempirical calculations.[32] Excellent results were reported with a mean unsigned error of only 1 kcal/mol for 399 neutral compounds.[32] Levy and co-workers have also developed highly accurate GBSA models, termed SGBNP[33] and AGBNP,[34] which employ OPLS-AA charges and radii and incorporate optimized nonpolar contributions to minimize errors with experimental results. Cramer, Truhlar, and co-workers have developed numerous solvation models, validated using much of the same experimental data used here, which have been subsequently incorporated into the AMSOL program.[6,35] Although highly accurate, AMSOL models tend to be highly parametrized, which makes incorporating routines for the computation of ΔG_{hyd} into a general molecular mechanics force field somewhat cumbersome. Many continuum models have been optimized to yield good ΔG_{hyd} results for small molecules using multiple atom types, radii, charges, and various combinations of adjustable nonpolar parameters which can limit transferability. Less-empirical models use very high-level ab initio and PB calculations with partial charges computed from SCRF methods with polarization[36] but are not easily adapted for general high-throughput screening.

Prompted by the need for a general continuum method with minimal optimized parameters, we have evaluated eight different point-charge models, based on ab initio, semiempirical, and empirical calculations in conjunction with a simple set of radii, through the computation of ΔG_{hyd} for small organic molecules. Computational results for more than 500 compounds (460 neutral compounds, 42 polyatomic ions, and 11 monatomic ions) are compared with experimental results, which, to our knowledge, represents the largest number of reference compounds employed for ΔG_{hyd} calculations. The evaluation of nonpolar contributions using an atom-based as opposed to a molecule-based solvent-accessible surface area term was also explored. Parameter set validation is critical since the use of different theoretical methods, atomic partial charge models, atomic radii, and nonpolar SA parameters will lead to different calculated ΔG_{hyd} results. The primary goal here was to assess the accuracy of the different charge models for neutral and charged compounds and to compare the results from both PBSA and GBSA continuum method calculations. The ultimate goal is to be able to easily incorporate accurate solvation effects using MM-GBSA methods for protein−ligand binding calculations and the rescoring of complexes obtained from high-throughput docking.

Computational Methods

Free Energies of Hydration (ΔG_{hyd}). As in prior PBSA and GBSA continuum studies,[25,26] the free energy of hydration is partitioned into two terms, polar and nonpolar, according to eq 1.

$$\Delta G_{hyd} = G_{polar} + G_{nonpolar} \quad (1)$$

Polar energies (G_{polar}) for PB calculations were obtained using a grid-based finite difference solution to the Poisson−Boltzmann equation with zero salt concentration (eq 2), where $\rho(r)$ is the charge distribution of the molecule, $\epsilon(r)$ is the dielectric constant, and $\phi(r)$ is the electrostatic potential. The solution of the PB equation for systems described by a classical force field yields the electrostatic potential at every grid point, and G_{polar} is then evaluated as a sum over all atoms (eq 3), where the partial atomic charge for atom i is multiplied by the difference in the computed grid-point potential ϕ_i for the transfer from the gas phase ($\epsilon = 1$) to water ($\epsilon = 80$).

$$\nabla[\epsilon(r)\,\nabla\phi(r)] = -4\pi\rho(r) \quad (2)$$

$$G_{polar} = \frac{1}{2}\sum_{i}^{N} q_i(\phi_i^{80} - \phi_i^{1}) \quad (3)$$

For GB calculations, G_{polar} contributions were obtained using eqs 4−5. Here, ϵ is the dielectric constant (80 for the water phase), q the partial atomic charges, r_{ij} the interatomic distance, and α_i are the Born radii, which are computed according to the pairwise descreening algorithm of Hawkins et al.[37,38]

$$G_{polar} = -166\left(1 - \frac{1}{\epsilon}\right)\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N}\frac{q_i q_j}{f_{GB}} - 166\left(1 - \frac{1}{\epsilon}\right)\sum_{i=1}^{N}\frac{q_i^2}{\alpha_i} \quad (4)$$

$$f_{GB} = \{r_{ij}^2 + \alpha_{ij}^2 \exp[-r_{ij}^2/(4\alpha_{ij}^2)]\}^{0.5} \quad (5)$$

Nonpolar contributions ($G_{nonpolar}$) to ΔG_{hyd} are estimated using only a simple solvent-accessible surface area term. Alternative procedures, which treat $G_{nonpolar}$ as two terms representing separate cavitation and solute−solvent dispersion contributions, have been recently reported by Gallicchio and Levy.[34] In the present paper, $G_{nonpolar}$ is estimated using either the total molecular SA (eq 6) or atomic-based SA_i (eq 7). Prior MM-PBSA and MM-GBSA binding energy protocols typically employed a molecular SA (eq 6) with $\gamma = 0.00542$ and $\beta = 0.92$, as recommended by Kollman and co-workers.[11,12] An alternative method, which was pursued in the present work, is to compute atom-based SA_i and optimize each SA constant using multiple linear regression to improve agreement with experiment (eq 7). Using atom-based SA_i contributions to estimate free energies of solvation was first proposed by Eisenberg and McLachlan,[39] and Scheraga and co-workers.[40]

$$G_{nonpolar} = (\gamma SA) + \beta \quad (6)$$

$$\Delta G_{hyd}(exptl) - G_{polar} = G_{nonpolar} = \sum_i \gamma_i SA_i \quad (7)$$

Table 1. Parameters for PB and GB Continuum Calculations

type	mbondi radii	Sx value (GB only)[a]	number of atoms[b]
HC	1.30[c]	0.85	4215
HN	1.30[c]	0.85	98
HO	0.80[c]	0.85	93
HS	0.80[c]	0.85	13
HP	1.30[c]	0.85	10
C	1.70[a]	0.72	2678
N	1.55[a]	0.79	128
O	1.50[a]	0.85	299
F	1.50[a]	0.88	53
P	1.85[a]	0.86	6
S	1.80[a]	0.96	26
Cl	1.70[a]	0.80	114
Br	1.85[d]	0.80	27
I	1.98[d]	0.80	12

[a] From AMBER version 7.[43] [b] See Supporting Information Table S1 for a listing of all compounds. [c] From refs 41 and 42. [d] From Bondi's original work.[47]

For a given set of calculations, PBSA or GBSA, the same structures, partial charges, and atomic radii were employed. Any differences in the final calculation results in this paper will, therefore, be only a function of the two different continuum theories.

Computation Details. A simple set of atomic radii based on the mbondi (modified bondi) scheme,[41,42] from the AMBER7 program,[43] was used in the calculations for neutral and polyatomic charged species. In the mbondi scheme, hydrogen atoms connected to carbon, sulfur, nitrogen, phosphorus, or oxygen (types HC, HS, HN, HP, or HO, respectively) can have unique radii (Table 1). Dielectric constants for all calculations (PB and GB) were set to 1, representing the gas phase, and 80, representing the water phase. PB calculations were performed using the program Delphi4[44,45] with the following parameters: boundary conditions = 4, internal dielectric constant = 1.0, external dielectric constant = 80.0, and scale = 4 grids/Å. Other Delphi parameters were assigned automatically using default values. Generalized Born calculations were performed using an in-house version of the Hawkins et al.[37,38] pairwise descreening model with scaling parameters (Sx values) adopted from Tsui and Case (Table 1).[41] The DMS program was used for all of the SA calculations.[46] In addition to the total SA value for a compound, DMS can be used to estimate atom-based surface areas (SA_i). For a given compound, the total solvent-accessible surface area should be equivalent to the sum of each atom-based solvent-accessible surface area ($SA = \sum SA_i$).

Molecular Structures and Experimental Data. Bordner et al.[31] have generously made available 410 neutral molecular structures along with the corresponding experimental log $P_{gas/water}$ partition coefficients from the tabulated work of Abraham et al.[3] (converted to free energies at 25 °C using $\Delta G_{hyd} = -2.3RT \log P_{gas/water}$). However, the Bordner set did not contain compounds with polar hydrogens connected to sulfur (HS; Table 1) or include charged species. We augmented the neutral set with 50 additional neutral compounds (including compounds containing HS), as well as 42

Table 2. Correlation Coefficients (r^2) and Average Unsigned Errors (aue) for Experimental[a] vs Calculated[b] (PBSA or GBSA) Free Energies of Hydration (ΔG_{hyd}) Using Standard SA Constants[c]

	neutral molecules, $N = 460$; part I				charged (± 1) molecules, $N = 42$; part II			
model	r^2 PBSA	aue	r^2 GBSA	aue	r^2 PBSA	aue	r^2 GBSA	aue
Gast	0.53	3.20	0.49	3.36	0.68	7.52	0.67	8.15
MMFF94	0.29	3.26	0.26	3.41	0.73	7.44	0.72	8.27
AM1BCC	0.74	1.36	0.70	1.38	0.56	8.28	0.53	9.64
AM1CM2	0.71	3.09	0.67	2.81	0.39	11.67	0.34	13.63
PM3CM2	0.69	2.79	0.64	2.61	0.62	10.84	0.62	11.90
MSK	0.77	1.54	0.72	1.63	0.74	6.42	0.72	7.30
RESP	0.77	1.47	0.72	1.51	0.75	6.34	0.73	7.20
ChelpG	0.73	1.61	0.69	1.67	0.74	6.36	0.72	7.28

[a] See Supporting Information Table S1 for experimental references. [b] Calculated values obtained using eq 1. G_{polar} from either PB or GB calculations. [c] $G_{nonpolar} = (0.00542 \times SA_{total}) + 0.92$. Energies in kcal/mol.

charged (± 1) polyatomic compounds and 11 ionic monatomic species (see Table S1, Supporting Information). All additional compounds were obtained from the NIST Chemistry WebBook database[48] or constructed using the MOE program.[49]

Partial Charge Models. Eight charge models were evaluated in this study: Gasteiger−Marsili (Gast),[50] MMFF94,[51] AM1BCC,[29,30] AM1CM2,[52] PM3CM2,[52] Merz−Singh−Kollman (MSK),[53] Restrained Electrostatic Potential (RESP),[54,55] and ChelpG.[56] While the preceding list is not exhaustive, it does include methods that are currently implemented in several molecular modeling packages and allow for the calculation of partial atomic charges for diverse organic molecules. For a comparison with the present work, Udier-Blagovic et al. have recently evaluated the accuracy of partial charges computed using CM1 and CM3 procedures.[28] The molecule database was maintained with the MOE program,[49] and several software packages were used to assign the different charge models. Gast and MMFF94 charges were assigned using MOE. AM1BCC charges were determined using the ANTECHAMBER module in AMBER7[43] from MOPAC[57] calculations. The AMSOL[35] program was used to compute AM1CM2 and PM3CM2 partial charges.[52] AMSOL calculations incorporated the SM5.42R[6] water solvent model, which allows the charges to be computed in a simulated condensed phase. The MSK, RESP, and ChelpG charges were computed at the HF/6-31G*//HF/6-31G* level of theory using the program Gaussian 98.[58] Molecules containing iodine used the 3-21G* basis set for iodine and 6-31G* for all other atoms. The ANTECHAMBER module in AMBER7 was used for two-stage RESP fittings. It should be mentioned that different software packages may yield slight variations in atomic charges because of differences in the implementation of a particular partial charge model. Only the above-named program implementations were evaluated in this report.

Molecule Geometries. For each compound, the partial charges obtained using the eight different methods were mapped back to one set of standard geometries. Using one set of conformations allows for a direct comparison of the accuracy of the partial charge models and removes the possibility that different geometries would affect the results. Here, the standard geometries were taken as those obtained from a gas-phase geometry optimization using the MMFF94 force field as implemented in the MOE program. In general, the optimizations yielded extended structures. Other geometries could have been used, although this was not explored. Given that the data set contains mostly rigid compounds, the effect of including multiple conformations on the computed free energies of hydration was not investigated. Of the 502 polyatomic compounds, more than half (53%) contain two or fewer rotatable bonds. Averaging over multiple conformations in the previous Bordner study changed the computed free energies by only a small amount.[31]

Results and Discussion

Charge Model Evaluation. Free energies of hydration were computed for comparison with experimental results for compounds employing one of eight partial charge models (Gast, MMFF94, AM1BCC, AM1CM2, PM3CM2, MSK, RESP, and ChelpG). Table 2 lists the correlation coefficients (r^2) and average unsigned errors (aue) between experiment and theory as obtained from PBSA and GBSA calculations. In Table 2, the $G_{nonpolar}$ term is computed from molecular SA (eq 6) using the standard MM-PBSA and MM-GBSA constants ($\gamma = 0.00542$ and $\beta = 0.92$). Results for charged and neutral compounds are always reported separately since artificially high r^2 squared values may result when correlations are computed using both species together. This is primarily due to the large difference in magnitude of the experimental data for charged versus neutral species.

The correlation coefficients for neutral compounds in Table 2 (part I) track with the eight different charge schemes in roughly the following order: ab initio (MSK, RESP, and ChelpG) > semiempirical (AM1BCC, AM1CM2, and PM3CM2) > empirical (Gast and MFF94). Ab initio charges yield PBSA and GBSA r^2 values from 0.69 to 0.77, semiempirical r^2 values from 0.64 to 0.74, and empirical r^2 values from 0.26 to 0.53. Average unsigned errors (aue) follow the r^2 trends; ab initio charges yield smaller errors (1.47−1.67 kcal/mol) than semiempirical (1.36−3.09 kcal/mol) or empirical (3.20−3.41 kcal/mol) charges. For comparison, results from various parametrizations of the AMSOL SM5.42R universal solvation models from Cramer, Truhlar, and co-workers yield small unsigned errors of 0.43−0.46 kcal/mol for 275 neutral solutes.[6] Gallicchio et al. have also reported errors of less than 0.5 kcal/mol using the optimized SGB/NP model.[33] The primary reason AMSOL and SGB/NP methods yield much smaller errors than the results

Table 3. Average Unsigned Errors (aue) for Experimental[a] vs GBSA Calculated[b] Free Energies of Hydration (ΔG_{hyd}) Using Standard SA Constants[c]

type	number	Gast	MMFF94	AM1BCC	AM1CM2	PM3CM2	MSK	RESP	ChelpG
alkanes	19	0.33	0.33	0.30	0.32	0.28	0.60	0.49	0.38
alkenes/dienes	11	0.83	0.83	0.23	0.22	0.14	1.23	1.21	0.70
alkynes	4	1.90	1.38	0.56	1.95	1.77	2.23	2.18	1.55
arenes	19	3.64	3.64	0.30	2.08	1.35	0.46	0.47	1.96
alcohols/phenols	26	4.66	4.66	0.88	1.07	0.85	1.96	1.77	1.59
ethers	15	2.20	2.20	2.07	1.24	1.72	1.87	1.97	2.03
ketones/aldehydes	20	2.87	2.87	2.73	6.13	5.59	1.48	1.37	1.33
carboxylic acids	3	4.32	4.32	3.14	10.61	10.46	7.01	6.93	7.29
esters	14	2.75	2.75	0.87	5.36	4.19	1.50	1.40	1.43
amines	21	5.86	5.86	2.88	2.97	3.47	2.89	2.97	3.02
amides	2	8.71	8.70	5.19	7.59	6.41	1.53	1.39	1.63
nitriles	4	5.95	5.95	1.14	8.18	8.92	4.39	3.76	4.26
nitrohydrocarbons	6	4.98	6.74	1.75	5.44	7.63	4.92	4.38	4.58
nitrogen heterocyclic	10	4.98	4.98	0.96	3.41	3.35	0.90	1.02	1.87
thiols	3	2.73	2.73	0.34	0.80	1.03	0.82	0.51	0.21
sulfides	3	3.43	3.43	1.32	0.76	0.85	0.65	1.04	1.46
all 180 molecules	180	3.43	3.48	1.37	2.85	2.75	1.76	1.69	1.85

[a] See Supporting Information Table S1 for experimental references. [b] Calculated values obtained using eq 1. G_{polar} from GB calculations. [c] $G_{nonpolar} = (0.00542 \times SA_{total}) + 0.92$. Energies in kcal/mol.

presented here in Table 2 is due to the fact that many parameters have been optimized to minimize errors with respect to experimental results. Similar approaches could be adopted by the present GBSA method through the incorporation of schemes which allow separate radii and nonpolar contributions to be optimized on the basis of unique atom types (e.g., aromatic vs aliphatic carbons). A very simple optimization, based on the elemental atom types listed in Table 1, is presented in a later section of this paper. In the present study, the best agreement with experiment, for all 460 neutral compounds, was obtained using AM1BCC charges, which yielded aue's of 1.36 and 1.38 kcal/mol from PBSA and GBSA calculations, respectively (Table 2). For comparison, in the recent Jorgensen et al. study,[32] a mean unsigned error of 1.01 kcal/mol was reported for GBSA results using OPLS-AA radii with scaled AM1CM1A charges for 399 neutral compounds. A smaller subset of 75 molecules in that report yielded a larger mean unsigned error of 1.51 kcal/mol, which dropped to 1.16 kcal/mol if nitro compounds and DMSO were excluded.[32]

Table 3 shows GBSA results obtained here, broken down by molecular class for 180 out of 199 neutral compounds in common with the Gallicchio et al. study, which employed the SGB/NP model.[33] As previously noted, the SGB/NP model is a fitted method and, therefore, yields aue errors much lower than those reported here (aue = 0.32 kcal/mol fitted, 0.50 kcal/mol jackknife).[33] However, Table 3 highlights the fact that, for many molecule classes, low errors can, in fact, be obtained using a very simplistic GBSA model not fit a priori to reproduce experimental results. In particular, excellent results are obtained with the AM1BCC model, which yields a low aue of only 1.37 kcal/mol. The three ab initio methods yield errors between 1.69 and 1.85 kcal/mol. Other models yield larger aue's between 2.75 and 3.48 kcal/mol. Notably, the AM1BCC charges yield errors of less than 1 kcal/mol for more than half the compounds tested in Table 3 and include alkanes, alkenes/dienes, alkynes, arenes, alcohols/phenols, esters, nitrogen heterocyclics, and thiols. The major outliers for AM1BCC are amides ($N = 2$), carboxylic acids ($N = 3$), amines ($N = 21$), and ketones/aldehydes ($N = 20$). The largest errors using ab initio charges in Table 3 are for carboxylic acids, amines, nitriles ($N = 4$), and nitrohydrocarbons ($N = 6$). The largest errors reported in the prior SGB/NP[33] study also included nitriles (jackknifed aue = 1.15 kcal/mol) and nitro compounds, (jackknifed aue = 2.57 kcal/mol), in addition to nitrogen heterocyclics (jackknifed aue = 1.22 kcal/mol), and indicate the challenges associated with obtaining accurate charge distributions for nitrogen-containing species.

It should be noted that, during parametrization of the AM1BCC method, adjustments were made to the way partial charges are computed so that calculated relative free energies of solvation for amines, nitros, and unsaturated hydrocarbons were in closer agreement to experimental results.[29,30] However, the AM1BCC method was not optimized for use with the GBSA model or mbondi radii utilized here. Given this fact, the results in Tables 2 and 3 are extremely encouraging given that low errors can be obtained for most molecules and that AM1BCC charges are extremely fast to generate for databases containing even hundreds of thousands of diverse molecules.[30]

Surprisingly, the three semiempirical methods tested here yield poorer correlations with experimental results than do the two empirical methods for charged (± 1) molecules (Table 2, part II). Ab initio charges yield the strongest correlations, with r^2 values from 0.72 to 0.75, compared to semiempirical r^2 values from 0.34 to 0.62 and empirical r^2 values from 0.72 to 0.73. Given that Jorgensen et al. obtained good results for 17 polyatomic charged compounds using the OPLS-AA GBSA models augmented with unscaled AM1CM1 charges,[32] the poor results obtained here using AM1CM2 and PM3CM2 are unexpected. AM1CM2 and PM3CM2 methods should yield partial charges qualitatively similar to those obtained from the AM1CM1 procedure. Results for 22 charged

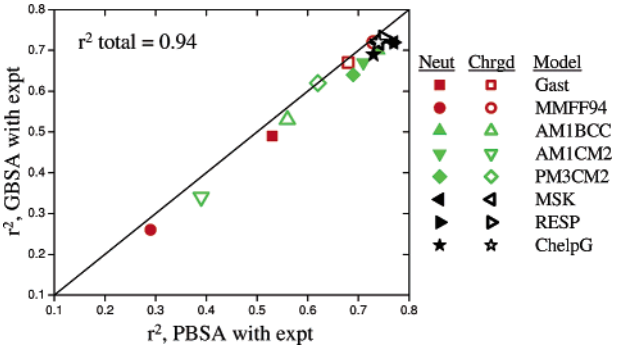


Figure 1. Comparison of correlation coefficients (r^2 values) for calculated versus experimental free energies of hydration from PBSA and GBSA calculations. For each partial charge model, two r^2 values are plotted representing results for 460 neutral compounds (filled symbols) and 42 charged compounds (open symbols)(see Table 2). The overall correlation between the total PBSA and GBSA results is $r^2 = 0.94$.

compounds using the more sophisticated SGB/NP model with OPLS-AA charges yielded jackknifed aue's of 4.12 and 2.23 kcal/mol for 3 carboxylate anions and 19 ammonium cations, respectively. The larger errors reported here for charged species compared with those of other studies could arise from differences such as the number and type of compounds tested, the functional form of the GB equation used to compute G_{polar}, the atomic radii used in the calculations, or a lack of fitted nonpolar contributions. As emphasized previously, only element-based radii, with the exception of hydrogen atoms connected to C, N, O, P, and S (Table 1), were used. Li et al. estimate that the typical uncertainty in the experimental data for ions is larger, at about 5 kcal/mol, in comparison with neutral compounds, which is typically 0.2 kcal/mol.[6]

As was the case for neutral compounds, the aue errors reported here (Table 2) track with the correlation coefficients. Again, ab initio partial charges yield the lowest errors (6.34−7.30 kcal/mol), but for the charged species, semiempirical charges yield the largest errors (8.28−13.63 kcal/mol). Empirical aue's are in the middle (7.44−8.27 kcal/mol). Thus, using MSK, RESP, and ChelpG, partial charges for neutral and charged species consistently yield the strongest correlations and lowest average unsigned errors with experimental free energies of hydration regardless of which continuum method was employed for the computation (Table 2). The r^2 values for three ab initio methods cluster around 0.75 for both neutral and charged species (Figure 1).

PBSA versus GBSA. The PBSA and GBSA results are highly correlated and independent of the charge model used for the calculations (Table 2; Figure 1). The strong agreement between PBSA and GBSA r^2 values (obtained from computed versus experimental results) suggests that a given partial charge model will influence the final free energies much more than which continuum method (PBSA or GBSA) is used for the calculations. Correlation coefficients between PB and GB polar energies are always very strong, $r^2 > 0.94$, and independent of which partial charge model or data set (neutral or charged compounds) was employed in the calculations. These trends continue to provide strong support for using GBSA methods as a reasonable alternative to the more computationally demanding PBSA calculations for free energy calculations.

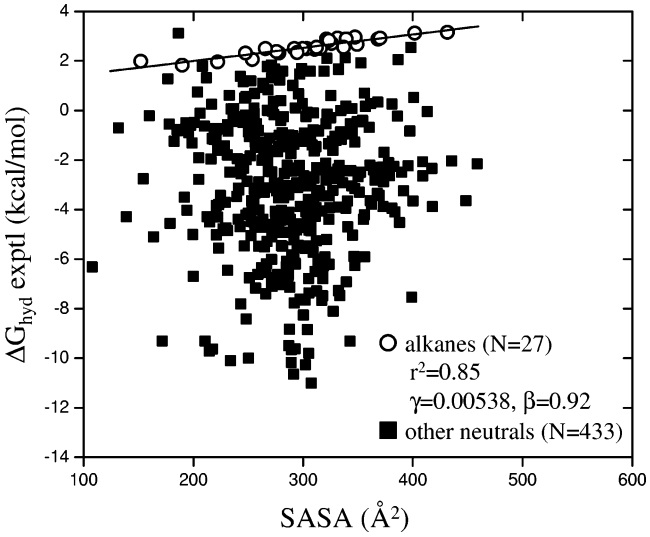


Figure 2. Experimental free energies of hydration vs total molecular solvent-accessible surface area (SASA). The best fit line to the 27 linear and branched alkanes (○) yields a correlation coefficient $r^2 = 0.85$, $\gamma = 0.005\,38$, and $\beta = 0.92$. Other compounds are represented as filled squares (■).

***$G_{nonpolar}$* from Molecular SA versus Atomic SA$_i$.** The constants ($\gamma = 0.00542$ and $\beta = 0.92$) typically used[13−20] in MM-PBSA and MM-GBSA calculations to convert SA (Å^2) to $G_{nonpolar}$ (kcal/mol) are based on fitting molecular SA results to experimental ΔG_{hyd} values for small straight-chain alkanes.[25] The rationale for this procedure exploits the fact that alkanes have low dipole moments and nonpolar contributions will, therefore, dominate ΔG_{hyd}. Figure 2 shows the molecular SA for the 460 neutral molecules studied here versus experimental ΔG_{hyd} values along with the best fit regression line using only the 27 linear and branched alkanes.

The constants obtained from this linear regression fit (Figure 2, open circles, $r^2 = 0.85$, $m = 0.00538$, $b = 0.92$) are essentially identical to the standard constants ($\gamma = 0.00542$ and $\beta = 0.92$).[13−20] However, as a group, molecular SAs have no correlation with experimental ΔG_{hyd} values (Figure 2, filled squares). Rankin et al. have reported similar results based on an analysis of 210 neutral compounds.[59] In most cases, G_{polar} contributions are the dominant factor for the final correlations with experimental results for the neutral molecules reported in Table 2. This is illustrated in Figure 3 for 460 neutral compounds in which the polar energies (G_{polar} $r^2 = 0.77$, filled squares) computed from PB calculations with RESP charges strongly correlate with experimental ΔG_{hyd} values. However, nonpolar contributions computed using standard SA constants (eq 6; $\gamma = 0.00542$ and $\beta = 0.92$) yield no correlation ($G_{nonpolar}$ $r^2 = 0.00$, open circles) and, for this data set, do not contribute to any improvement or diminishment in the final correlation coefficient with experimental results (ΔG_{hyd} $r^2 = 0.77$).

To improve the agreement with experimental results, we explored a procedure first proposed by Eisenberg and McLachlan[39] and Scheraga and co-workers,[40] which optimizes nonpolar contributions using atom-based coefficients for a given SA type through multiple linear regression

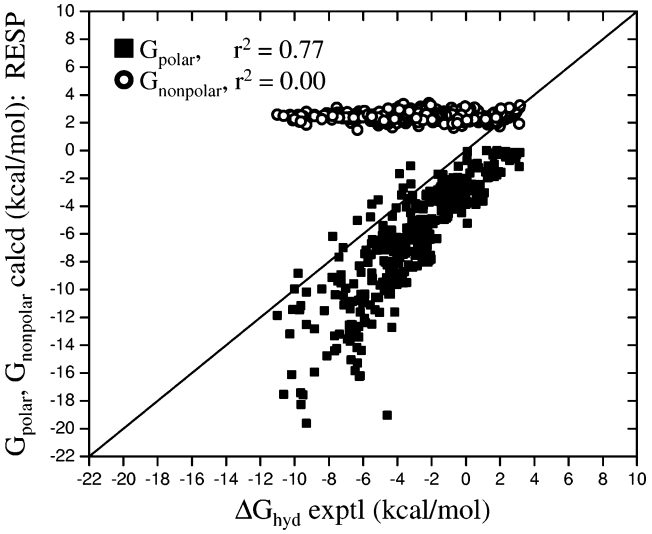


Figure 3. Correlation of individual components with experimental free energies of hydration for neutral compounds ($N = 460$) using RESP-derived partial charges. Polar (■) energies G_{polar} from PB calculations. Nonpolar (○) energies from molecular solvent-accessible surface area calculations $G_{nonpolar} = (0.00542 \times SA_{total}) + 0.92$.

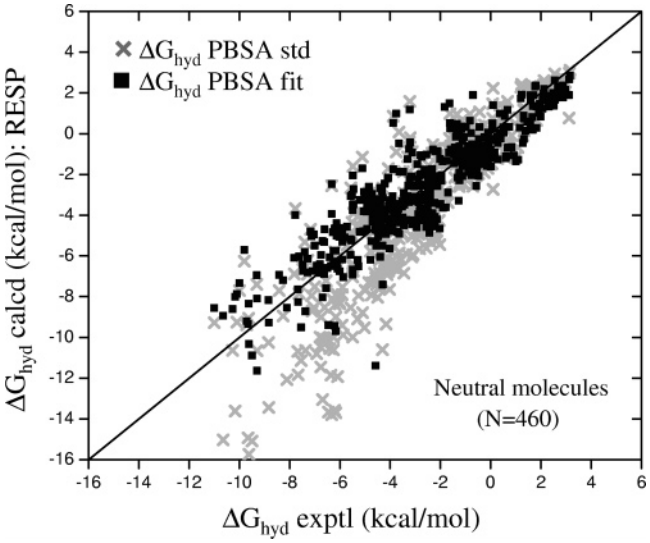


Figure 4. Predicted free energies of hydration (ΔG_{hyd} calcd) vs experimental free energies of hydration (ΔG_{hyd} exptl) from PBSA calculations with RESP charges for neutral compounds ($N = 460$). Nonpolar energies from molecule-based SAs using standard constants (×) or atom-based SAs using fitted constants (■).

fitting.[39,40] In the present work, SA γ_i coefficients were optimized for each mbondi type (HC, HN, HO, HS, HP, C, N, O, F, P, S, Cl, Br, and I) using multiple linear regression (eq 7) by fitting to the residuals in exptl $\Delta G_{hyd} - \Delta G_{polar}$ computed using the eight different charge models from either GB or PB calculations. After the fittings, new $G_{nonpolar}$ contributions were recomputed using the atom-based constants (γ_i's) so that optimized ΔG_{hyd} values could then be compared with experimental results. Levy and co-workers have reported an alternative functional form for $G_{nonpolar}$ in which atom-based coefficients are used to compute separate cavitation and solute−solvent dispersion interactions, as implemented in the SGB/NP[33] and AGBNP[34] models.

In most cases, utilizing the new SA_i constants to estimate nonpolar energies improves the overall agreement with experimental ΔG_{hyd} values (Table 4 versus Table 2). In particular, overall aue's are substantially reduced and r^2 values improve in general. However, using fitted constants actually reduces correlations (r^2) for neutral compounds for the three semiempirical models (AM1BCC, AM1CM2, and PM3CM2) despite the fact that the aue's show dramatic improvement (Table 4 versus Table 2). Here, degradations in r^2, for neutral molecules with semiempirical charges, arise because the fitting procedure attempts to minimize the overall error (neutral and charged) with experimental results, and these models originally performed poorly for charged species. The most robust and best overall improvement is obtained for the ab initio methods that consistently yielded the best agreement with experimental results for both neutral and charged species.

Figures 4 and 5 highlight favorable cases where the use of atom-based constants yield improved results even if a particular charge model already leads to good agreement with experimental results. Ab initio charges (MSK, RESP, and ChelpG) yield G_{polar} energies in strong correlation with those of the experiment for neutral and charged compounds in all cases. However, using molecule-based constants (gray crosses) to compute $G_{nonpolar}$ values can lead to a systematic overestimate (absolute error) of the hydration free energies for species with ab initio charges in the experimental range from −11 to −2 kcal/mol for neutrals and −90 to −60 kcal/mol for charged species. As an example, for neutrals, Figure

Table 4. Correlation Coefficients (r^2) and Average Unsigned Errors (aue) for Experimental[a] vs Calculated[b] (PBSA or GBSA fit) Free Energies of Hydration (ΔG_{hyd}) Using Fitted SA Constants[c]

	neutral molecules, $N = 460$; part I				charged (±1) molecules, $N = 42$; part II			
model	fitted r^2 PBSA	aue	fitted r^2 GBSA	aue	fitted r^2 PBSA	aue	fitted r^2 GBSA	aue
Gast	0.67	1.43	0.56	1.62	0.69	8.60	0.69	8.99
MMFF94	0.36	1.91	0.28	2.07	0.70	8.24	0.68	8.60
AM1BCC	0.68	1.26	0.58	1.49	0.61	6.71	0.60	6.83
AM1CM2	0.62	1.71	0.54	1.83	0.55	7.35	0.58	7.55
PM3CM2	0.61	1.66	0.52	1.83	0.68	7.24	0.71	7.47
MSK	0.81	0.99	0.69	1.32	0.79	4.46	0.77	4.68
RESP	0.80	1.02	0.69	1.33	0.80	4.45	0.78	4.69
ChelpG	0.81	0.99	0.70	1.30	0.79	4.46	0.77	4.67

[a] See Supporting Information for experimental references. [b] Calculated values obtained using eq 1. G_{polar} from either PB or GB calculations. [c] $G_{nonpolar} = \sum \gamma_i SA_i$. Energies in kcal/mol.

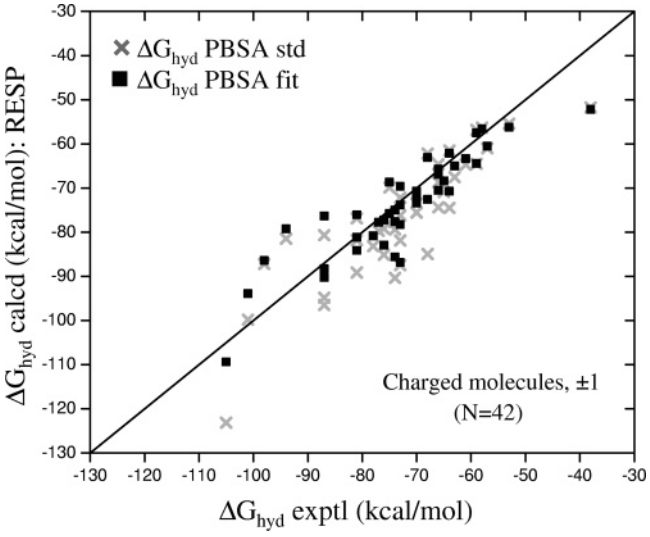


Figure 5. Predicted free energies of hydration (ΔG_{hyd} calcd) vs experimental free energies of hydration (ΔG_{hyd} exptl) from PBSA calculations with RESP charges for charged compounds ($N = 42$). Nonpolar energies from molecule-based SAs using standard constants (×) or atom-based SAs using fitted constants (■).

4 shows that using γ_i constants optimized from PBSA-RESP fits leads to an improvement of r^2 from 0.77 (ΔG_{hyd} std, gray crosses) to 0.80 (ΔG_{hyd} fit, black squares), and the aue error with the experiment drops from 1.47 to 1.02 kcal/mol (Figure 4). More dramatic results are observed for charged species; PBSA correlations increase from 0.75 (ΔG_{hyd} std, gray crosses) to 0.80 (ΔG_{hyd} fit, black squares), and the aue error with the experiment drops dramatically from 6.34 to 4.45 kcal/mol (Figure 5).

The primary motivation for using atom-based SA_i instead of molecule-based SA procedures is to reduce errors with respect to experimental results in three ways: (1) remedy gross deficiencies a particular charge model may have (r^2 and aue), (2) fine-tune an already reasonable agreement with experimental results (primarily aue), or (3) account for nonoptimal radii. On a case-by-case basis, simple atom-based constants (Figures 4−5, black squares) can correct for systematic errors. However, additional improvement, beyond what is presented in Tables 2, 3, and 4, would probably require changes to the model to include more atom-typing, adjustable radii, and nonpolar parameter optimization as in other methods.[6,33,34]

Optimized SA Coefficients. Tables 5 and 6 list "sets" of optimized SA_i constants (γ_i) obtained from multiple linear regressions using PB and GB G_{polar} results for all eight charge models employed in the calculations. For a new calculation that employs a particular charge model, atom-based γ_i values can be used to estimate $G_{nonpolar}$ energies that should lead to improved ΔG_{hyd} calculations. Despite the fact that G_{polar} results from both continuum methods show strong correlation (Table 2; Figure 1), for completeness, separate fits were performed for PB- (Table 5) or GB-derived (Table 6) G_{polar} energies.

As averaged over the entire data set of 502 molecules, the magnitude and sign for each γ_i value can give some indication as to the error with the experiment (and direction) associated with a particular charge model for a given atom (mbondi) type. However, caution should be exercised when trying to ascribe too much physical significance to any given SA coefficient. For some atom types listed in Table 1, HS ($N = 13$), P ($N = 6$), and I ($N = 12$), a lack of experimental data could potentially lead to SA optimizations that are underdetermined. Nevertheless, given the fact that related charge methods such as AM1CM2/PM3CM2 or MSK/RESP yield similar fitted SA constants (Tables 5 and 6), the multiple linear regression results appear robust. As an example, phosphorus (mbondi type P) coefficients from GB fits for AM1CM2 and PM3CM2 charged compounds are relatively large in magnitude compared with other types (Tables 5 and 6). Here, the negative coefficients are always in the range −1.8 to −2.8. The large negative sign indicates that, on average, G_{polar} terms computed using AM1CM2 and PM3CM2 charges underestimate the experimental ΔG_{hyd} values. Nonpolar contributions computed using atom-based SA_i will yield a favorable free energy to correct for this underestimation given that SA is always a positive value and, in this case, the γ_i for P atoms are negative. On the other hand, the GB γ_i coefficients for atom types P for ab initio-based methods (MSK, RESP, and ChelpG) are positive and

Table 5. Optimized Atomic SA Coefficients (γ_i Values)[a] Obtained Using Poisson−Boltzmann (PB) Derived G_{polar} Energies

type	Gast	MMFF94	AM1BCC	AM1CM2	PM3CM2	MSK	RESP	ChelpG
Hc	0.00093	−0.00002	0.00355	0.00962	0.00827	0.00679	0.00687	0.00649
Ho	−0.00434	−0.11172	0.25999	0.12379	0.11210	0.37414	0.36422	0.36037
Hs	0.28952	0.23307	0.33731	−0.53475	−0.45896	0.05493	0.06772	0.09424
Hn	−0.04103	−0.02779	−0.01058	−0.01094	−0.01857	−0.00574	−0.00436	−0.00813
Hp	−0.12342	0.00990	0.02589	0.47729	0.38605	−0.02415	−0.00164	−0.01025
C	−0.01634	−0.01610	0.02001	0.04395	0.03708	0.01765	0.01468	−0.00278
N	−0.00798	−0.01032	0.07251	0.05061	0.08398	0.04518	0.04440	0.05156
O	0.00759	0.04621	0.02409	0.09277	0.08863	0.03592	0.03292	0.04072
F	0.02036	0.02024	0.02256	0.02661	0.01954	0.01755	0.01643	0.01873
P	2.12323	0.36337	0.98863	−2.44577	−2.59507	0.92016	0.61608	0.79176
S	0.01477	0.02908	0.05082	0.15426	0.13041	0.04414	0.04145	0.03315
Cl	0.00336	0.00302	0.00384	0.00330	0.00662	0.00560	0.00527	0.00657
Br	−0.00532	−0.00455	0.00139	−0.00410	0.00415	0.00681	0.00550	0.00479
I	−0.00635	−0.00609	0.01495	−0.01134	−0.00775	0.00656	0.00562	−0.00116

[a] $G_{nonpolar} = \sum \gamma_i SA_i$ optimized using neutral ($N = 460$) and charged ($N = 42$) compounds.

Table 6. Optimized Atomic SA Coefficients (γ_i Values)[a] Obtained Using Generalized Born (GB) Derived G_{polar} Energies

type	Gast	MMFF94	AM1BCC	AM1CM2	PM3CM2	MSK	RESP	ChelpG
Hc	−0.00049	−0.00146	0.00129	0.00719	0.00588	0.00489	0.00484	0.00436
Ho	0.02765	−0.07826	0.25937	0.13669	0.12418	0.36583	0.35732	0.35663
Hs	0.29006	0.23719	0.30314	−0.48392	−0.40299	0.07870	0.09415	0.13314
Hn	−0.04816	−0.02950	−0.02299	−0.02113	−0.02386	−0.01374	−0.01218	−0.01520
Hp	−0.07575	0.06191	0.07913	0.54968	0.46106	0.01841	0.03414	0.02993
C	−0.01537	−0.01529	0.01715	0.03967	0.03379	0.02164	0.01859	0.00328
N	0.01065	0.00709	0.10707	0.07294	0.10361	0.06938	0.06810	0.07333
O	0.01100	0.04920	0.02952	0.09624	0.09423	0.04269	0.03965	0.04760
F	0.02353	0.02559	0.02941	0.02826	0.02085	0.02082	0.01948	0.02374
P	1.50762	−0.30940	0.71401	−1.75879	−2.78904	0.47251	0.25635	0.40528
S	0.01889	0.03237	0.05437	0.15530	0.13185	0.04452	0.04131	0.03165
Cl	0.00536	0.00489	0.00662	0.00515	0.00913	0.00878	0.00784	0.00787
Br	−0.00329	−0.00275	0.00466	−0.00130	0.00710	0.01492	0.01301	0.00786
I	−0.00419	−0.00384	0.02054	−0.00733	−0.00400	0.01865	0.01703	0.00294

[a] $G_{nonpolar} = \sum\gamma_i SA_i$ optimized using neutral ($N = 460$) and charged ($N = 42$) compounds.

Table 7. GBSA Results for Monoatomic Ions

ion	ΔG_{hyd} exptl[a]	OPLS-AA radii[b]	ABS error[c]	adjusted radii	ABS error[c]
F^-	−107	1.540	4.54		
Cl^-	−78	2.090	2.21		
Br^-	−72	2.255	1.89		
I^-	−63	2.700	2.26		
Li^+	−122	1.350	6.66	1.370[d]	4.62
Na^+	−98	1.680	3.53		
K^+	−81	2.020	2.22		
Mg^{2+}	−456	1.455	22.89	1.515[d]	2.64
Ca^{2+}	−381	1.735	16.01	1.785[d]	4.23
Fe^{2+}	−456			1.515[e]	2.64
Zn^{2+}	−485			1.435[e]	1.04

[a] See Supporting Information Table S1 for experimental references. [b] From Jorgensen et al.[32] [c] Absolute error for ΔG_{hyd} exptl − ΔG_{hyd} calcd; calculated values obtained using eq 1 with $G_{nonpolar} = (0.00542 \times SA_{total}) + 0.92$. [d] Adjusted from reference 32. [e] This work. Energies in kcal/mol.

much smaller at about 0.26−0.47. The variation in the optimized coefficients in Tables 5 and 6 is a direct result of the differences that are obtained from the different partial charge methods used for the computation of G_{polar}. Because of this fact, optimized constants can be viewed as a SA-based correction factor to account for errors in any particular charge model in an average sense. Moreover, γ_i constants should only be used in conjunction with the partial charge model with which they were derived.

Monatomic Ions. We have also pursued free energy of hydration calculations for 11 monatomic ions using the same GBSA protocols for comparison with experimental results. Monatomic ions are a unique case given that only a single atom is present, and therefore, they are not charge-model-dependent; only the formal ion charge and radius needs to be specified. Radii for most monatomic ions were taken from Jorgensen et al.[32] and used as reported or adjusted slightly for use with the present model. In general, nonpolar contributions to the total ΔG_{hyd} for monatomic ions would be negligible given the large polarization energy (−63 to −485 kcal/mol; Table 7) compared to the small solvent-accessible surface area contribution. The solvent-accessible surface area for a monatomic species is simply $SA = 4\pi(r + 1.4)^2$, where 1.4 Å represents the standard probe radius for water and r is the radius.

In general, the absolute difference between GBSA computed and experimental values (ΔG_{hyd} exptl − ΔG_{hyd} calcd) is lower than the estimated uncertainty for ions[6] (5 kcal/mol) using OPLS-AA radii with standard MM-GBSA constants ($G_{nonpolar}$ $\gamma = 0.00542$ and $\beta = 0.92$), as shown in Table 7. However, for Li^+, Mg^{2+}, and Ca^{2+}, larger errors of 6.66, 22.89, and 16.01 kcal/mol, respectively, are observed. It should be noted that results from Jorgensen et al. for the monatomics agree exactly with experimental results.[32] The results presented here in Table 7 employed different SA constants than those used by Jorgensen et al.; thus, slight differences are not unexpected. For consistency and to optimize parameters for the present GBSA model, the radii for three ions, Li^+, Mg^{2+}, and Ca^{2+}, were adjusted slightly so that all errors for monatomics would be less than 5 kcal/mol. Since SA contributions to hydration free energies for monatomic species are assumed to be small, the dominant change from the adjustment of radii will be to the G_{polar} term. For ions in particular, continuum results are very sensitive to the choice of atomic radii. For example, GB results for the ±1 monatomic ions shown in Table 7 change by more than 7 kcal/mol with only a 0.1 Å change in radius (1.5−1.6 Å). The same change for ±2 monatomics changes GB results dramatically by more than 30 kcal/mol. Which radii to employ for PBSA and GBSA continuum calculations is the subject of considerable research.[25,60−62]

Conclusion

The primary goal of this study was to evaluate procedures for the computation of free energies of hydration, in the context of a general classical molecular mechanics force field, for use in the simulation of protein−ligand binding and virtual screening (docking). Improved computational procedures continue to advance the utility of structure-based drug design. Here, absolute free energies of hydration have been computed using continuum PBSA and GBSA methods for comparison with experimental results for a diverse set of 460 neutral compounds, 42 polyatomic ions, and 11 monatomic ions. A systematic evaluation of eight different models has revealed that continuum results for small organic

molecules with partial charges based on one of three ab initio methods consistently lead to the best overall correlation with experimental results for both neutral and charged species (Table 2; Figure 1). Correlation coefficients with the experiment using MSK, RESP, and ChelpG charges with GBSA yield r^2 values between 0.69 and 0.73 and with PBSA yield r^2 values between 0.72 and 0.77. The semiempirical AM1BCC model yields good results for neutral compounds with an r^2 value of 0.70−0.74 and the lowest aue's of all the models tested (aue = 1.36−1.38). However, the use of semiempirical (AM1BCC, AM1CM2, and PM3CM2) and empirical (Gast and MFF94) charge schemes yielded mixed results dependent on whether the compounds were charged or neutral (Table 2).

The computational results presented here clearly show that correlations with experimental ΔG_{hyd} values are independent of which implicit solvation model (PBSA or GBSA) is employed in the calculations. In all cases, the Hawkins pairwise GB results are strongly correlated (overall $r^2 = 0.94$) with the much more expensive PB calculations, provided that identical coordinates, radii, and atomic charges are used (Figure 1).

An examination of polar and nonpolar energy components shows that $G_{nonpolar}$ energies derived from molecule-based SAs and standard conversion constants have no correlation with experimental results for neutral compounds (Figure 3). The lack of a universal SA constant stems from the erroneous assumption that all exposed atoms contribute equally to nonpolar energies. In the present work, improved correlations with experimental results were obtained through simple optimizations of atom-based SA constants using multiple linear regression fits to the difference in experimental free energies and polar energy terms obtained from continuum calculations (Tables 4−6). On a case-by-case basis, using atom-based SA_i instead of molecule-based SA constants significantly reduces both relative (r^2) and absolute unsigned errors (aue) with respect to the experiment by eliminating any gross deficiencies a particular charge model may have (Tables 4 versus Table 2). In particular, aue's are substantially reduced (Table 4), and systematic errors can be corrected (Figures 4−5, black squares). As was the case using standard SA constants, the best agreement with experimental results using atom-based SA_i constants was obtained with ab initio partial charge models; improved r^2 values are between 0.69 and 0.78 and 0.79−0.81 from GBSA and PBSA calculations, respectively (Table 4).

Finally, studies that continue to assess the accuracy of atomic partial charges and other force field parameters are critically important for the field of computational structural biology. The results here show both the strengths and the weaknesses of various "Amber-like" approaches to force fields for docking and screening. Hence, the results will be of interest to those considering such calculations. The fact that hydration energies calculated by other models (e.g., AMSOL or SGB/NP) show better agreement with experimental results than those reported here is useful information that may help researchers in their choice of computational strategies. The primary motivation here was to determine generally useful force-field parameters for estimating changes in the free energy of hydration associated with molecular recognition for use with MM-PBSA and MM-GBSA and for docking calculations. On the basis of a comparison with experimental ΔG_{hyd} values for more than 500 diverse organic molecules, MSK, RESP, or ChelpG partial atomic charges obtained from ab initio calculations using 6-31G* wave functions would be recommended for both charged and neutral species if computational resources allow it. For molecule libraries requiring partial charge assignments for hundreds of thousands of compounds, the semiempirical AM1BCC method would be recommended over the more approximate alternatives tested.

Acknowledgment. Gratitude is expressed to Walter Mangel for helpful discussions and to the Department of Defense (Award Number DAMD17-00-1-0192, Modification P00001) and National Institutes of Health (GM-56531, P. Ortiz deMontellano, P. I.) for support of this research. T.A. was supported by a U. C. president's dissertation year fellowship.

Supporting Information Available: Experimental free energies of hydration (Table S1). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Leo, A.; Hansch, C.; Elkins, D. Partition Coefficients and Their Uses. *Chem. Rev.* **1971**, *71*, 525−616.

(2) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Delivery Rev.* **2001**, *46*, 3−26.

(3) Abraham, M. H.; Whiting, G. S.; Fuchs, R.; Chambers, E. J. Thermodynamics of Solute Transfer from Water to Hexadecane. *J. Chem. Soc. Perkin Trans. 2* **1990**, 291−300.

(4) Chambers, C. C.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Model for aqueous solvation based on class IV atomic charges and first solvation shell effects. *J. Phys. Chem.* **1996**, *100*, 16385−16398.

(5) Gerber, P. R. Charge distribution from a simple molecular orbital type calculation and nonbonding interaction terms in the force field MAB. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 37−51.

(6) Li, J. B.; Zhu, T. H.; Hawkins, G. D.; Winget, P.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. Extension of the platform of applicability of the SM5.42R universal solvation model. *Theor. Chem. Acc.* **1999**, *103*, 9−63.

(7) Åvist, J. Ion−Water Interaction Potentials Derived form Free Energy Perturbation Simulations. *J. Phys. Chem.* **1990**, *94*, 8021−8024.

(8) Babu, C. S.; Lim, C. Theory of ionic hydration: Insights from molecular dynamics simulations and experiment. *J. Phys. Chem. B* **1999**, *103*, 7958−7968.

(9) Marcus, Y. A Simple Empirical Model Describing the Thermodynamics of Hydration of Ions of Widely Varying Charges, Sizes, and Shapes. *Biophys. Chem.* **1994**, *51*, 111−127.

(10) Srinivasan, J.; Cheatham, T. E.; Cieplak, P.; Kollman, P. A.; Case, D. A. Continuum solvent studies of the stability

of DNA, RNA, and phosphoramidate−DNA helices. *J. Am. Chem. Soc.* **1998**, *120*, 9401−9409.

(11) 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*, 889−897.

(12) 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.

(13) Kuhn, B.; Kollman, P. A. Binding of a diverse set of ligands to avidin and streptavidin: An accurate quantitative prediction of their relative affinities by a combination of molecular mechanics and continuum solvent models. *J. Med. Chem.* **2000**, *43*, 3786−3791.

(14) Wang, J.; Morin, P.; Wang, W.; Kollman, P. A. Use of MM-PBSA in reproducing the binding free energies to HIV-1 RT of TIBO derivatives and predicting the binding mode to HIV-1 RT of efavirenz by docking and MM-PBSA. *J. Am. Chem. Soc.* **2001**, *123*, 5221−5230.

(15) 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.

(16) Huo, S.; Wang, J.; Cieplak, P.; Kollman, P. A.; Kuntz, I. D. Molecular dynamics and free energy analyses of cathepsin D-inhibitor interactions: insight into structure-based ligand design. *J. Med. Chem.* **2002**, *45*, 1412−1419.

(17) Wang, W.; Lim, W. A.; Jakalian, A.; Wang, J.; Luo, R.; Bayly, C. I.; Kollman, P. A. An analysis of the interactions between the Sem-5 SH3 domain and its ligands using molecular dynamics, free energy calculations, and sequence analysis. *J. Am. Chem. Soc.* **2001**, *123*, 3986−3994.

(18) Suenaga, A.; Hatakeyama, M.; Ichikawa, M.; Yu, X.; Futatsugi, N.; Narumi, T.; Fukui, K.; Terada, T.; Taiji, M.; Shirouzu, M.; Yokoyama, S.; Konagaya, A. Molecular dynamics, free energy, and SPR analyses of the interactions between the SH2 domain of Grb2 and ErbB phosphotyrosyl peptides. *Biochemistry* **2003**, *42*, 5195−5200.

(19) Donini, O. A. T.; Kollman, P. A. Calculation and prediction of binding free energies for the matrix metalloproteinases. *J. Med. Chem.* **2000**, *43*, 4180−4188.

(20) 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.

(21) Jorgensen, W. L. Free Energy Calculations: A Breakthrough For Modeling Organic Chemistry in Solution. *Acc. Chem. Res.* **1989**, *22*, 184−189.

(22) Kollman, P. Free Energy Calculations: Applications to Chemical and Biochemical Phenomena. *Chem. Rev.* **1993**, *93*, 2395−2417.

(23) Jorgensen, W. L.; Ravimohan, C. Monte Carlo Simulation of Differences in Free Energies of Hydration. *J. Chem. Phys.* **1985**, *83*, 3050−3054.

(24) Cramer, C. J.; Truhlar, D. G. Implicit Solvation Models: Equilibria, Structure, Spectra, and Dynamics. *Chem. Rev.* **1999**, *99*, 2161−2200.

(25) Sitkoff, D.; Sharp, K. A.; Honig, B. Accurate Calculation of Hydration Free-Energies Using Macroscopic Solvent Models. *J. Phys. Chem.* **1994**, *98*, 1978−1988.

(26) 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.

(27) 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*, 265−284.

(28) Udier-Blagovic, M.; Morales De Tirado, P.; Pearlman, S. A.; Jorgensen, W. L. Accuracy of free energies of hydration using CM1 and CM3 atomic charges. *J. Comput. Chem.* **2004**, *25*, 1322−1332.

(29) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method. *J. Comput. Chem.* **2000**, *21*, 132−146.

(30) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parametrization and validation. *J. Comput. Chem.* **2002**, *23*, 1623−1641.

(31) Bordner, A. J.; Cavasotto, C. N.; Abagyan, R. A. Accurate Transferable Model for Water, n-Octanol, and n-Hexadecane Solvation Free Energies. *J. Phys. Chem. B* **2002**, *106*, 11009−11015.

(32) Jorgensen, W. L.; Ulmschneider, J. P.; Tirado-Rives, J. Free Energies of Hydration from a Generalized Born Model and an All-Atom Force Field. *J. Phys. Chem. B* **2004**, *108*, 16264−16270.

(33) Gallicchio, E.; Zhang, L. Y.; Levy, R. M. The SGB/NP hydration free energy model based on the surface generalized born solvent reaction field and novel nonpolar hydration free energy estimators. *J. Comput. Chem.* **2002**, *23*, 517−529.

(34) Gallicchio, E.; Levy, R. M. AGBNP: an analytic implicit solvent model suitable for molecular dynamics simulations and high-resolution modeling. *J. Comput. Chem.* **2004**, *25*, 479−499.

(35) Hawkins, G. D.; Giesen, D. J.; Lynch, G. C.; Chambers, C. C.; Rossi, I.; Storer, J. W.; Li, J.; Winget, P.; Rinaldi, D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *AMSOL*, version 6.6; University of Minnesota: Minneapolis, Minnesota. Based in part on *AMPAC*, version 2.1, by Liotard, D. A.; Healy, E. F.; Ruiz, J. M.; Dewar, M. J. S. and on *EF* by Jensen, F.

(36) Marten, B.; Kim, K.; Cortis, C.; Friesner, R. A.; Murphy, R. B.; Ringnalda, M. N.; Sitkoff, D.; Honig, B. New model for calculation of solvation free energies: Correction of self-consistent reaction field continuum dielectric theory for short-range hydrogen-bonding effects. *J. Phys. Chem.* **1996**, *100*, 11775−11788.

(37) 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.

(38) 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.

(39) Eisenberg, D.; McLachlan, A. D. Solvation Energy in Protein Folding and Binding. *Nature* **1986**, *319*, 199−203.

(40) Ooi, T.; Oobatake, M.; Nemethy, G.; Scheraga, H. A. Accessible Surface-Areas as a Measure of the Thermodynamic Parameters of Hydration of Peptides. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 3086−3090.

(41) 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.

(42) Tsui, V.; Case, D. A. Theory and applications of the generalized Born solvation model in macromolecular simulations. *Biopolymers* **2000**, *56*, 275−291.

(43) 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*, 1668−1688.

(44) Rocchia, W.; Alexov, E.; Honig, B. Extending the applicability of the nonlinear Poisson−Boltzmann equation: Multiple dielectric constants and multivalent ions. *J. Phys. Chem. B* **2001**, *105*, 6507−6514.

(45) Rocchia, W.; Sridharan, S.; Nicholls, A.; Alexov, E.; Chiabrera, A.; Honig, B. Rapid grid-based construction of the molecular surface and the use of induced surface charge to calculate reaction field energies: applications to the molecular systems and geometric objects. *J. Comput. Chem.* **2002**, *23*, 128−137.

(46) *DMS*; University of California at San Francisco Computer Graphics Laboratory: San Francisco, CA.

(47) Bondi, A. van der Waals Volumes and Radii. *J. Phys. Chem.* **1964**, *68*, 441−451.

(48) NCI Public Database. http://dtp.nci.nih.gov.

(49) *MOE*, version 2002.03; Chemical Computing Group: Montreal, Canada.

(50) Gasteiger, J.; Marsili, M. Iterative Partial Equalization of Orbital Electronegativity − a Rapid Access to Atomic Charges. *Tetrahedron* **1980**, *36*, 3219−3228.

(51) Halgren, T. A. Merck molecular force field. 1. Basis, form, scope, parametrization, and performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490−519.

(52) Li, J. B.; Zhu, T. H.; Cramer, C. J.; Truhlar, D. G. New class IV charge model for extracting accurate partial charges from wave functions. *J. Phys. Chem. A* **1998**, *102*, 1820−1831.

(53) Besler, B. H.; Merz, K. M.; Kollman, P. A. Atomic Charges Derived from Semiempirical Methods. *J. Comput. Chem.* **1990**, *11*, 431−439.

(54) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using-Charge Restraints For Deriving Atomic Charges − the RESP Model. *J. Phys. Chem.* **1993**, *97*, 10269−10280.

(55) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollman, P. A. Application of Resp Charges to Calculate Conformational Energies, Hydrogen-Bond Energies, and Free-Energies of Solvation. *J. Am. Chem. Soc.* **1993**, *115*, 9620−9631.

(56) 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.

(57) Stewart, J. J. P.; Rossi, I.; Hu, W.-P.; Lynch, G. C.; Liu, Y.-P.; Chuang, Y.-Y.; Li, J.; Cramer, C. J.; Fast, P. L.; Truhlar, D. G. *MOPAC*, version 5.09mn; University of Minnesota: Minneapolis, Minnesota.

(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.

(59) Rankin, K. N.; Sulea, T.; Purisima, E. O. On the transferability of hydration-parametrized continuum electrostatics models to solvated binding calculations. *J. Comput. Chem.* **2003**, *24*, 954−962.

(60) Onufriev, A.; Case, D. A.; Bashford, D. Effective Born radii in the generalized Born approximation: the importance of being perfect. *J. Comput. Chem.* **2002**, *23*, 1297−1304.

(61) Nina, M.; Beglov, D.; Roux, B. Atomic radii for continuum electrostatics calculations based on molecular dynamics free energy simulations. *J. Phys. Chem. B* **1997**, *101*, 5239−5248.

(62) Banavali, N. K.; Roux, B. Atomic radii for continuum electrostatics calculations on nucleic acids. *J. Phys. Chem. B* **2002**, *106*, 11026−11035.

CT050097L

JCTC Journal of Chemical Theory and Computation

Force Field Effects on a β-Sheet Protein Domain Structure in Thermal Unfolding Simulations

Ting Wang* and Rebecca C. Wade

Molecular and Cellular Modeling Group, EML Research, Schloss-Wolfsbrunnenweg 33, 69118 Heidelberg, Germany

Received June 22, 2005

Abstract: The secondary structure propensities observed in protein simulations depend heavily on the force field parameters used. The existing empirical force fields often have difficulty in balancing the relative stabilities of helical and extended conformations. The resultant secondary structure bias may not be apparent in short simulations at room temperature starting from the native folded states. However, it can manifest itself dramatically at high temperatures and lead to large deviations from experimentally observed secondary structure propensities. Motivated by thermal unfolding simulations of several WW domains, which have a three-stranded β-sheet structure, we chose the FBP28 WW domain as a well-characterized system to investigate several AMBER force fields as well as parametrization of the NPSA (Neutralized, Polarized ionizable side chains with a solvent-accessible Surface Area-dependent term) implicit solvent model. The ff94 force field and two variants with altered parameters for the backbone torsion term were found to convert the native β-sheet structure directly to a single helix at high temperatures, whereas the ff96 force field produced significant non-native β-sheet content at high temperatures. The ff03 force field was able to reproduce the β-sheet-coil transition and experimentally observed unfolding pathways with both an explicit water solvent and the NPSA implicit solvent model at relatively low temperatures. However, the protein domain became predominantly helical after unfolding. Modification of the solvation parameter in the NPSA implicit solvent model was not sufficient to remedy this problem. The results imply that the intrinsic secondary structure bias in a force field cannot easily be solved by modifying a single parameter such as backbone torsion potential or a solvation parameter of a solvent model. Nevertheless, the results show that the AMBER ff03 force field together with an explicit solvent model or the NPSA implicit solvent model is a useful tool for studying the unfolding of both α- and β-sheet structure protein domains, and an integrative consideration of all force field parameters is likely to be necessary for a complete solution.

Introduction

Empirical molecular mechanics force field parameters are generally used for macromolecular modeling and simulation due to the unaffordable computational cost of performing ab initio quantum mechanics calculations. Widely used molecular mechanics force fields include AMBER,[1−4] CHARMM,[5,6] GROMOS,[7] and OPLS.[8] These force fields produce reasonable results for many studies. However, because empirical force fields are parametrized by fitting to experimental results and ab initio quantum mechanics calculations for a limited number of small peptides or nucleotides, difficulties exist when the parameters are applied to proteins and nucleic acids. For this reason, force fields are continually being improved on the basis of the increasing understanding of proteins and nucleic acids and advances in computational methodology and computer power. For example, the AMBER force field has developed from ff94,[1] ff96,[2] and ff99[3] to the current ff03.[4] In simulations of

* Corresponding author e-mail: ting.wang@eml-r.villa-bosch.de.

10.1021/ct0501607 CCC: $33.50
Published on Web 12/08/2005

dynamic processes such as protein folding, unfolding, and flexible docking, the accurate treatment of protein conformations and interactions is crucial. This requires accurate parameters for both helical conformations and extended conformations. Helical conformations are strongly dependent on local interactions, e.g. i, i+4 hydrogen-bonding interactions in the α-helix, whereas β-sheet conformations are more influenced by nonlocal (in sequence) interactions. Thus, it is more difficult to model β-sheets. In addition, most of the model peptide systems used to derive force field parameters, such as polyalanine peptides, are systems that preferentially sample helix-coil conformational space. As a consequence, the existing force fields often show a bias toward overstabilizing α-helical and understabilizing β-extended conformations.[2,4,9−15] Many parameters can affect protein conformations, including atomic charges, nonbonded interaction parameters, and backbone torsion (ϕ and φ) angle parameters. Among these, backbone torsion (ϕ and φ) angle parameters are the most directly related to protein secondary structure formation and therefore often used to adjust the secondary structure propensity of a force field. A number of force field evaluations have been conducted by Garcia's group[10,13] and Pande's group.[11,16] The model peptides studied were helix-coil transition systems, and helical propensity was the main concern. A short 12-residue β-hairpin tryptophan zipper was studied by Simmerling and co-workers[15] to evaluate the AMBER ff94 and ff99 force fields, and the peptide was found to convert to a stable α-helix at 550 K with both ff94 and ff99 force fields. Here, we evaluate the AMBER force fields by simulation of a 3-stranded β-sheet protein domain to investigate the relative stability and the balance between helical and extended conformations. As far as we know, this investigation of the secondary structural conformational preferences of different force fields is the first based on a protein system that undergoes β-sheet-coil transitions.

This paper was initiated by our study of the relative stability of WW domains and their mutants by thermal unfolding simulations carried out with AMBER force fields and our NPSA (Neutralized, Polarized ionizable side chains with a solvent-accessible Surface Area-dependent term) implicit solvent model.[17] WW domains are small 3-stranded β-sheet protein domains with two signature tryptophan (W) residues. Extensive experiments have shown that WW domains undergo a β-sheet-coil transition upon unfolding.[18−22] However, in our early simulations with the AMBER ff94 force field, we observed that the β-strands converted into α-helices during high temperature simulations, independent of the WW domain sequence. This result prompted us to investigate the available AMBER force fields and compare their performance in terms of unfolding behavior. Because of this motivation, most simulations in this study are conducted at high temperature to enable unfolding to occur on computationally accessible time scale. Most of simulations are conducted for the FBP28 WW domain, a WW domain with an experimentally well-defined structure and comparatively high thermal stability. The folding/unfolding of this domain has been extensively studied by NMR and CD spectroscopy.[18−22] The third strand was observed to be less stable than the first two strands and to be lost first upon unfolding. The folding/unfolding of FBP28 was observed to be three-state.

Another motivation for this paper is concerned with solvent models. Previous studies revealing the helix-favoring bias in the AMBER ff94 and ff99 force fields and the extended-favoring bias in the AMBER ff96 force field were done using an explicit solvent model[10,11,13,16,23,24] or the generalized Born implicit solvent model.[9,23] In this paper, we also investigate use of our NPSA implicit solvent model,[17] which is computationally advantageous compared to both the generalized Born model and an explicit solvent model and gives good results compared to other implicit solvent models in simulations of a variety of proteins at 300 K.[17] Therefore, the evaluation of the ff94 and ff96 force fields was conducted with the NPSA implicit solvent model only. For the newer ff03 force field, we performed simulations with both the NPSA implicit solvent model and an explicit solvent model.

Materials and Methods

The FBP28 WW domain folds into a twisted three-stranded antiparallel β-sheet structure with a melting temperature of 64 °C (337 K).[22] In this study, we used the first structure in the NMR ensemble (PDB entry 1e0l) of FBP28 WW domain. It has 37 residues.

NPSA Model. The simulations of FBP28 were carried out by using the AMBER7 program, modified to incorporate the NPSA implicit solvent model. The NPSA model has been demonstrated to be efficient for maintaining protein native structures and flexibly docking proteins at 300 K.[17] The model uses a distance-dependent dielectric function ($\epsilon = r$), but the partial charges of the ionized side-chains (residues Glu, Asp, Lys, and Arg) and the N- and C-termini are neutralized (N) and polarized (P), and a solvent-accessible surface area (SA)-dependent term ($\Delta G = \sum \sigma A$) is added. The modified charges (called NPSA charges) were parametrized based on the AMBER ff94 force field. The solvation parameter σ was set as 0.012 kcal/mol·Å^2 for nonpolar carbon and sulfur atoms and −0.06 kcal/mol·Å^2 for polar nitrogen and oxygen atoms.

For the ff03 force field, we reparametrized the NPSA charges for use with the new atomic partial charges in ff03 (see Table 1). With ff03, two different values of the solvation parameter σ for backbone nitrogen atoms, −0.06 and −0.12 kcal/mol·Å^2, were tested and compared while retaining 0.012 kcal/mol·Å^2 for nonpolar carbon and sulfur atoms and −0.06 kcal/mol·Å^2 for polar oxygen atoms and side chain nitrogen atoms.

Backbone Torsion Parameters. In addition to the standard ff94, we have studied four variants with altered backbone torsion parameters: ff96,[2] Simmerling's parameters,[9] Garcia's parameters,[10] and the ff03 ϕ/φ parameters. The backbone torsion potentials (C−N−CA−C (ϕ) and N−CA−C−N (φ)) studied are plotted in Figure 1, except for Garcia's parameters in which the backbone torsion potential is set to zero. Peptide ϕ/φ dihedral angle energy terms are generally represented by a Fourier series of cosine functions that contribute additively in a force field. Modifications take place in the coefficients and phases of the cosine functions. From the plots in Figure 1, we can see that in

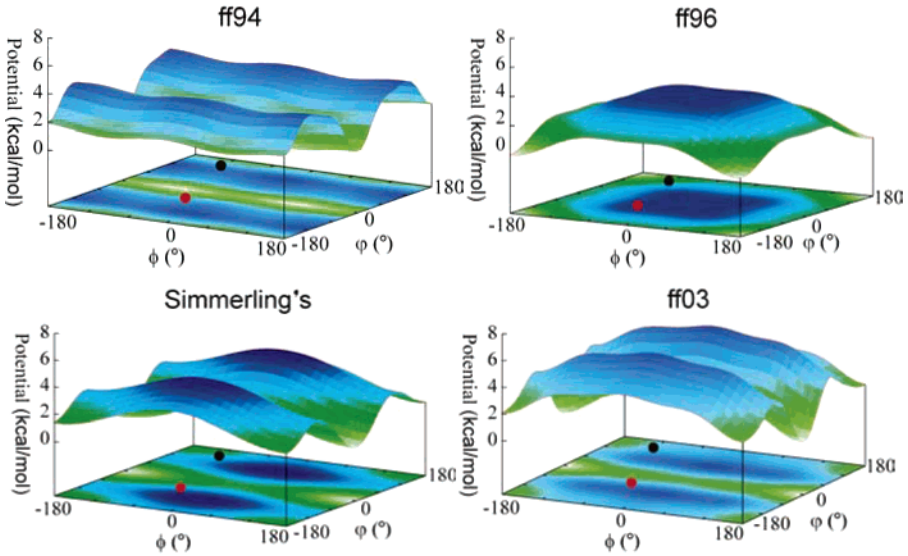


Figure 1. The backbone torsion potentials (C−N−CA−C (ϕ) and N−CA−C−N (φ)) studied in this paper, include ff94,[1] ff96,[2] Simmerling's,[9] Garcia's,[10] and the ff03[4] force field. The backbone torsion potential is set to zero in Garcia's modification and is therefore not plotted in this figure. The red dots indicate the helical conformations ((ϕ,φ) = (−60°, −40°)), and the black dots indicate β-extended conformations ((ϕ,φ) = (−120°, 140°)). In ff94, the helical regions are closer to the energy minimum than extended regions, whereas in ff96, helical regions lie on the energy maximum. In Simmerling's modification and ff03, both helical and extended conformations are located intermediates between the energy minimum and maxima.

Table 1. Modified Partial Atomic Charges (e) for Ionized Side Chains and the N- and C-Termini Used in the NPSA Model Based on the ff03 AMBER Force Field[c]

	ASP		GLU		LYS		ARG	
ionizable side chains	CG	1.3452	CD	1.3652	NZ	−1.4504	NH1,2	−0.7858
	OD1,2	−0.5804	OE1,2	−0.6740	HZ1,2,3	0.3946	HH1,2,3,4	0.3411
	CB	0.0519	CG	0.0659	CE	−0.1698	CZ	0.0655
							HE	0.2262
NTER[a]	H = original charge + 0.1							
	N = original charge − 1.3							
CTER[b]	O = original charge + 0.2							
	C = original charge + 0.6							

[a,b] In the AMBER force field, the charges of the side chain atoms of ASP, GLU, LYS, and ARG are slightly different when they are terminal residues. The modified charges of the side chains of these terminal residues are therefore slightly different from those listed above but were assigned by following the same logic as in the reference paper.[17] The full list of NPSA partial atomic charges is available in the Supporting Information. [c] The side chains are neutralized and polarized in the NPSA model.

ff94, helical regions are closer to the energy minimum than extended regions, whereas in ff96, helical regions lie on the energy maximum. In both Simmerling's modification and ff03, helical and extended regions are located intermediate between the energy minimum and maxima.

Simulation Protocol. In simulations with the NPSA implicit solvent model, the structures of FBP28 were first energy minimized for 200 steps and then gradually heated from 0 K to the desired temperatures in 50 ps. They were then simulated at that temperature with a temperature-coupling constant of 1.0 ps. The bonds involving hydrogen atoms were constrained by using the SHAKE algorithm. A time step of 2 fs was used, and the nonbonded interactions were updated every 10 time steps with a cutoff of 10 Å.

In simulations with an explicit water model, the structure of FBP28 was solvated in a truncated octahedron TIP3P water box with a minimal distance of 12.0 Å between the boundaries of the box and the nearest protein atoms. 4391 water molecules were added to the system. The water molecules were subjected to 2000 steps of energy minimization, and then the whole system was energy minimized for 1000 steps. After energy minimization, the whole system was subjected to a gradual heating from 0 K to the desired temperature in 50 ps and kept at that temperature with a temperature coupling constant of 1.0 ps. The bonds involving hydrogen atoms were constrained by using the SHAKE algorithm. The time step was 2 fs, and the nonbonded interactions were updated every 10 time steps with a cutoff of 10 Å. The Particle Mesh Ewald (PME) method was used for the long-range electrostatic interactions with the default parameters. Constant volume was used in the first 300 ps, and a constant pressure of 1.0 atm was used for the rest of the simulation time.

The secondary structure element assignment was calculated with the program STRIDE[25] embedded in the software VMD 1.8.3.[26]

Results and Discussion

ff94, ff96. FBP28 was first simulated with the ff94 force field and the original NPSA parameters at 430 K. The

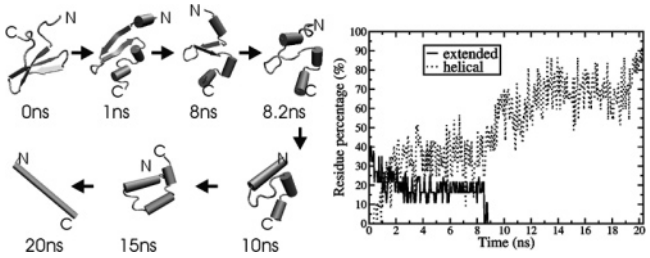


Figure 2. The 3-stranded β-sheet native structure of FBP28 gradually converted to a single helix when simulated at 430 K with the NPSA model and the ff94 force field. The right-hand plot shows the time development of the percentage of residues in extended (solid line) and helical (dash line) conformations. Significant helices appeared after ca. 1 ns and continued growing to occupy the whole protein. There was not even transient loss of helical content.

simulation temperature 430 K is higher than the melting temperature of FBP28 (337 K[18,22]), and the native 3-stranded β-sheet structure was indeed lost. However, the protein did not change to a random coil structure but gradually and directly converted to a single helix in 20 ns as shown in Figure 2. The plot in Figure 2 shows the time development of the percentage of residues in extended (solid line) and helical (dash line) conformations. Significant helices appeared after ca. 1 ns and continued increasing to occupy the whole protein. There was not even a transient coil state between the initial β-sheet structure and the final α-helix structure. This result is in disagreement with experiments[18−22] where no structured protein was observed in the unfolded states. The helical content of FBP28 was predicted to be 0.75% and 0% by the program Agadir[27] and the Web server PredictProtein,[28] respectively. The overstabilization of helix is apparent.

With respect to the unfolding pathway, the first two strands were much more persistent than the third strand, and this yielded a constant percentage of residues in extended conformations between 1 ns and 8 ns in the plot in Figure 2. This is in agreement with the order of loss of β-strands observed in unfolding experiments. Importantly, regardless of the helical content, the long time persistence of the first two strands implies an intermediate between folded and unfolded states, which is consistent with the 3-state unfolding behavior observed in experiments[22] and previous unfolding simulations of FBP28 WW domain with explicit water solvent.[21] Nevertheless, the substantial helical content prevented the observation of a realistic unfolding pathway.

Several studies have reported that, by modifying backbone torsion parameters, α-helix propensity can be reduced to approach experimental values.[3,9,10,15] These studies were however based on α-helix-coil systems not β-sheet-coil systems. We tried two of the backbone torsion parameter variants[9,10,15] suggested by Simmerling and co-workers[9] and by Garcia and co-workers,[10] respectively. The backbone torsion potential with Simmerling's parameters can be seen in Figure 1; both helical and extended regions lie between the energy minimum and maxima. Garcia's modification entails simply zeroing out the backbone torsion potential. Figure 3 shows the time development of the percentage of residues in extended (left) and helical (right) conformations with different backbone torsion parameters (ϕ/φ) applied to the ff94 force field together with the original NPSA implicit solvent model at 430 K. The 3-stranded β-sheet native structure ultimately converted to single helices in all simulations although with different conversion speeds. The simulations with the four different backbone torsion parameters yielded very similar profiles of the time development of the percentage of residues in extended and helical conformations. This result indicates that, in the context of the ff94 force field, these modifications of backbone torsion parameters alone cannot rectify the overstabilization of helical conformations.

However, the situation changed dramatically when we switched to the ff96 force field, which differs from ff94 only in the backbone torsion parameters. They were reparametrized to improve the stability of β-extended structures.[2] In Figure 1, we can see that helical regions lie on the energy maximum in the backbone torsion potential in the ff96 force field. In the simulations with ff96, the 3-stranded β-sheet of FBP28 was very stable, and complete unfolding occurred only at a highly elevated temperature of 600 K. Figure 4

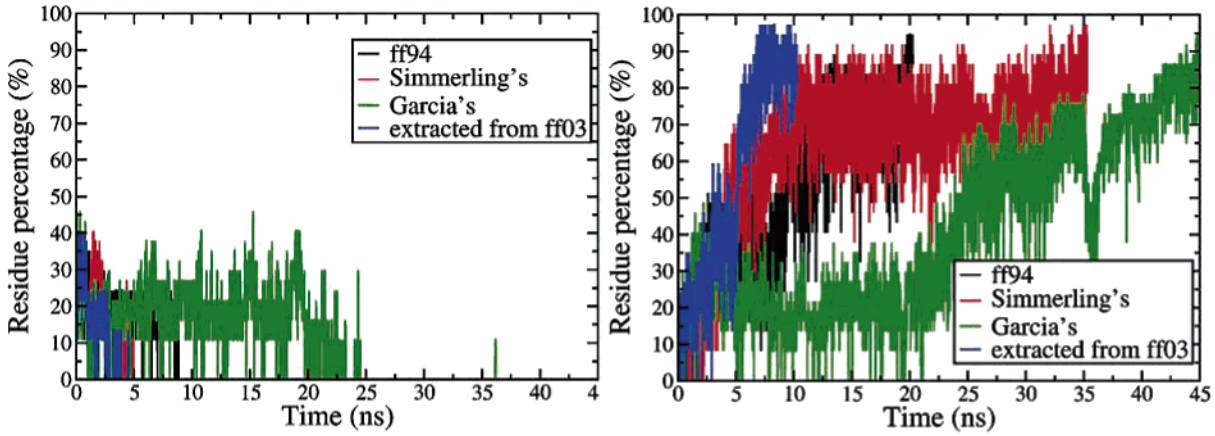


Figure 3. Time development of the percentage of residues in extended (left) and helical (right) conformations in the simulations of FBP28 with different backbone torsion parameters (ϕ/φ) applied to the ff94 force field together with the NPSA implicit solvent model at 430 K. black: ff94; red: Simmerling's; green: Garcia's; blue: (ϕ/φ) parameters extracted from ff03. The 3-stranded β-sheet native structure ultimately converted to single helices in all simulations although with different conversion speeds.

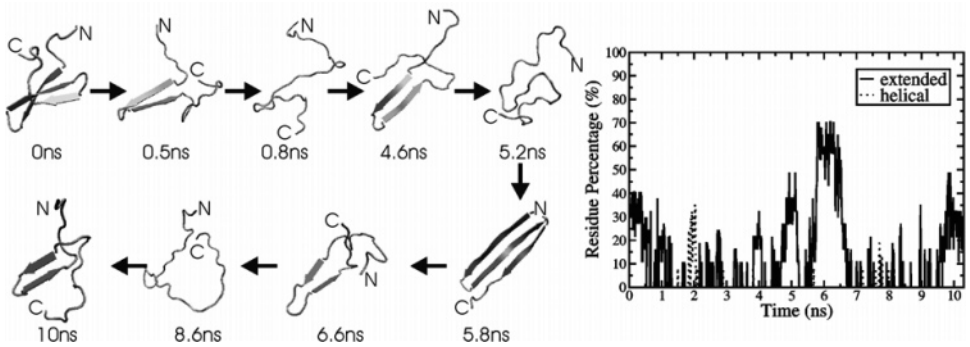


Figure 4. Representative structures of FBP28 in a 10-ns simulation with the NPSA model and the ff96 force field at 600 K. The right-hand plot shows the time development of the percentage of residues in extended (solid line) and helical (dash line) conformations. The 3-stranded β-sheet native structure unfolded in the first 1.3 ns and afterward non-native β-sheet structures appeared as the main populations. From 5.8 ns to 6.5 ns the whole structure became a long 3-stranded β-sheet. Only negligible transient helical conformations appeared.

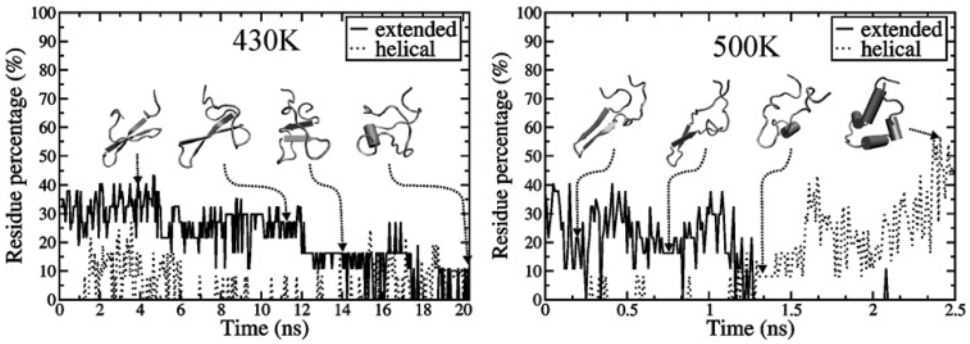


Figure 5. Time development of the percentage of residues in extended (solid line) and helical (dash line) conformations in the simulations of FBP28 with the NPSA model and the ff03 force field at 430 K (left) and 500 K (right). Although notable helical conformations are present in the trajectories, they are discontinuous until unfolding completed, losing all strands. In the left plot at 430 K, the time development of the percentage of extended conformations exhibits three steps: (1) the full the native 3-stranded structure (ca. 0−5 ns), (2) the full maintenance of the first two strands and the absence of the third strand (ca. 5−12 ns), and (3) the partial maintenance of the first two strands (ca. 12−18 ns). This enabled the observation of a clear unfolding pathway: the early loss of the third strand followed by the loss of the first two strands.

shows the representative structures of FBP28 in a 10-ns simulation at 600 K and the time development of the percentage of residues in extended (solid line) and helical (dash line) conformations. The 3-stranded β-sheet native structure unfolded in the first 1.3 ns, and afterward non-native β-sheet structures appeared as the main populations. From 5.8 ns to 6.5 ns, the whole structure became a long 3-stranded β-sheet. Only negligible transient helical conformations appeared. This result indicates the overstabilization of extended conformations in the ff96 force field, as observed in other studies using an explicit solvent model.[16,24] Although very short, the first 1.3 ns of the trajectory of FBP28 exhibited the unfolding pathway observed in experiments,[21] that is the early loss of the third strand followed by the loss of the first two strands.

ff03. With the recent distribution of the AMBER8.0 program, the ff03 force field[4] became available. The differences between ff03 and ff94 are in the atomic partial charges and backbone torsion parameters. The backbone torsion potential plot in Figure 1 shows that both helical and extended regions lie intermediate between the energy minimum and maxima.

We reparametrized our NPSA charges to be consistent with the new partial atomic charges. See Table 1 for the new NPSA charges. We conducted simulations for FBP28 at three different temperatures: 370 K, 430 K, and 500 K. At each temperature, three 20-ns runs were performed with different heating speeds to the desired temperature. At 370 K, the three-stranded β-sheet structure was stable and conserved throughout the three 20-ns simulations.

At 430 K, the structure partially unfolded during one of the three 20-ns simulations with the third strand swinging out first and then the first two strands separating. Complete unfolding occurred in the other two 20-ns simulations. The left-hand plot in Figure 5 shows the time development of the percentage of residues in extended (solid line) and helical (dash line) conformations in one of the two complete unfolding simulations at 430 K. Although notable helical conformations were still present in the trajectories, they were discontinuous until unfolding completed and all strands were lost. The time development of the percentage of residues in extended conformations exhibits three steps: (1) the full native 3-stranded structure (ca. 0−5 ns), (2) the full maintenance of the first two strands with the absence of the third strand (ca. 5−12 ns), and (3) the partial maintenance of the first two strands (ca. 12−18 ns). The last two steps together imply an intermediate between the folded and unfolded states, which is consistent with the 3-state unfolding

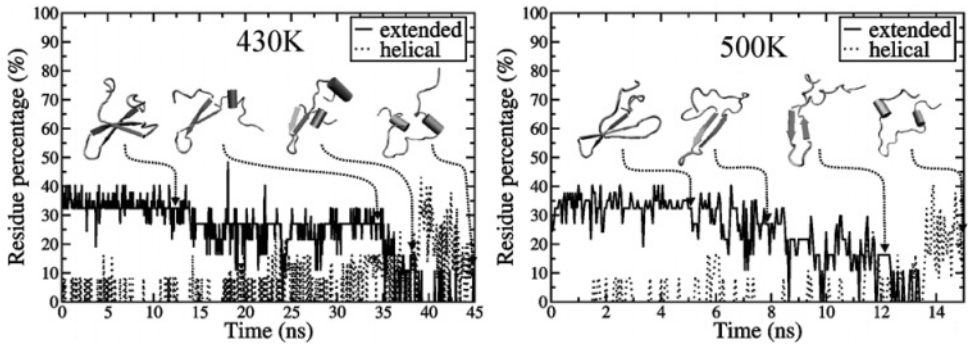


Figure 6. Time development of the percentage of residues in extended (solid line) and helical (dash line) conformations in the simulations of FBP28 with an explicit water solvent and the ff03 force field at 430 K (left) and 500 K (right). Similar to Figure 5 from the implicit NPSA solvent model, notable helical conformations are present in the trajectories but discontinuous until after complete unfolding. After unfolding, helices are the main secondary structure population. Another similarity between Figures 6 and 5 is that the unfolding pathway is clearer at 430 K than at 500 K because of the shorter unfolding time and significant fluctuation of the percentage of extended residues at 500 K.

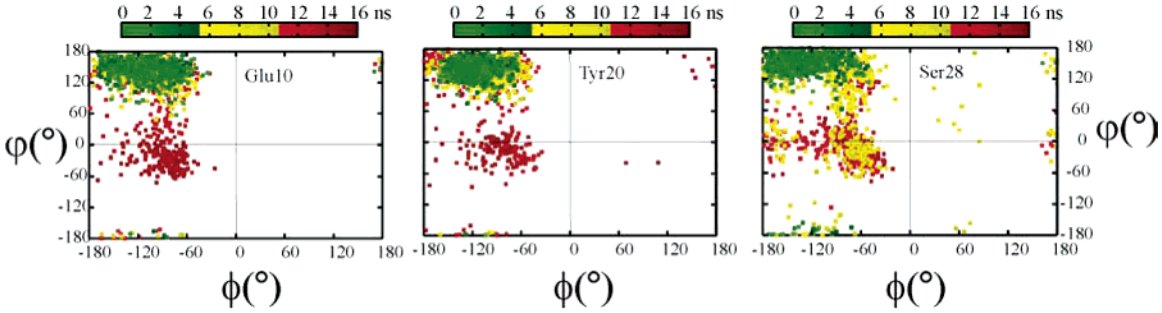


Figure 7. ϕ/φ distributions of the Glu10, Tyr20, and Ser28 residues of FBP28 in the simulation with explicit water solvent and the ff03 force field at 500 K. Glu10, Tyr20, and Ser28 are the middle residues of the first, the second, and the third strands in the native structure, respectively. The colors show the simulation time: beginning (green) to end (red).

behavior of FBP28 observed in experiments[22] and previous MD unfolding simulations with an explicit solvent model.[21] The transient nature of the helical content enabled the observation of a clear unfolding pathway: the persistence of the first two strands with the absence of the third strand and then the absence of all the strands in a predominant coiled structure. These results indicate that the ff03 force field has lower helix preference than the ff94 force field and a lower β-extended preference than the ff96 force field. The ff03 force field indeed has a better balance between α-helical and β-extended conformations than both ff94 and ff96, as stated in its reference paper.[4] We believe that the better balance comes from the new atomic charges as replacing the torsion parameters in ff94 with those in ff03 alone did not achieve such a result (see the blue lines in Figure 3).

At 500 K, unfolding was much quicker than at 430 K. The unfolding pathway was blurred by the significant fluctuation of the extended residues, and the helical content increased to more than 50% at ca. 2.4 ns (see the right-hand plot in Figure 5).

Despite the significant improvement of ff03 over ff94 and ff96, the overbiasing toward helical conformations still exists. To ensure that the effect of the implicit solvent model did not result in those biases, we conducted simulations with explicit water molecules for FBP28 at 430 K and 500 K. Figure 6 shows the time development of the percentage of residues in extended (solid line) and helical (dash line) conformations in a 45-ns trajectory at 430 K and a 16-ns trajectory at 500 K. The profiles are very similar to those with the NPSA model shown in Figure 5. Notable helical conformations are present in the trajectories but discontinuous until complete unfolding. After unfolding, helical conformations became the main population. Another similarity between simulations with the NPSA implicit solvent model and the explicit solvent model (Figures 6 and 5) is that the unfolding pathway is clearer at 430 K than at 500 K because of the shorter unfolding time and significant extended residue fluctuation at 500 K. Figure 7 shows the time development of the ϕ/φ distributions of the middle residues of the three native strands: Glu10, Tyr20, and Ser28, respectively, in the 500 K simulation. It is evident that all three residues converted from extended starting conformations to helical conformations at the end. The conversion of Ser28 in the third strand occurs earlier than that of Glu10 and Tyr20 in the first and the second strands. This is consistent with the lower stability of the third strand.

For comparison, we also conducted the same simulations for another WW domain, the YAP65 WW domain, which is less thermostable than FBP28.[18] The force field parameters in ff94, ff96, and ff03 showed similar effects for YAP65 and FBP28, although YAP65 exhibited less thermal stability by unfolding at lower temperatures (data not shown).

In addition, we simulated a 20-residue helical protein, the trp-cage miniprotein (PDB entry 1L2Y), with the NPSA implicit solvent model and the ff03 force field. The trp-cage protein folds into an α-helix and a short 3−10 helix with a

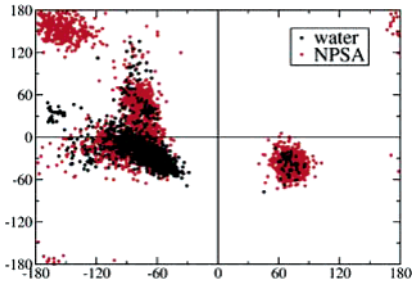


Figure 8. ϕ/φ distributions of the C-terminal residues Gln34, Glu35, and Leu36 of FBP28 in the simulations with the ff03 force field at 300 K. Black dots: with explicit water solvent; red dots: with $\sigma_{N\_bone} = -0.12$ kcal/mol·Å^2 in the NPSA model. The data for each model are from three 20-ns trajectories with different heating speeds to 300 K.

melting temperature of 42 degrees (315 K).[29] Our simulations were carried out at 5 different temperatures of 300 K, 315 K, 340 K, 370 K, and 500 K for 20 ns. At 300 K, both the secondary structure and the hydrophobic contacts between the residues in the trp cage were well maintained. Unfolding occurred on this time scale when the temperature was increased to 500 K, with the unstructured C-terminal segment separating from the helical N-terminal segment accompanied by the early loss of the short 3−10 helix. These results indicate that the reparametrized NPSA charges for the ff03 force field can maintain the protein native structures at room temperature and reproduce experimentally observed loss of structure at high temperatures for both β-sheet proteins and helical proteins.

We have investigated the effect of backbone torsion parameters in the context of the ff94 force field and the effect of the combination of backbone torsion parameters and atomic partial charges in the ff03 force field. The results demonstrated the superiority of the latter in terms of improving the balance between helical and extended conformations. Recently, Pande and co-workers[11] reported a result from the interplay between turning off the torsion potential, 1−4 charge−charge interactions and 1−4 van der Waals interactions for a helix-coil transition system. They found that the effects of these factors are complex being force-field dependent and nonadditive. This also implies that the improvement of empirical force fields is complex, and alteration of one parameter may require the modification of other parameters.

Solvation Parameter σ in the NPSA Model. Although there is no evidence to show that water plays a specific structural role in the folding/unfolding process of WW domains, the solvent properties and the treatment of solvent effects can affect the free energy landscape of folding/unfolding. Helical and extended conformations show different extents of solvent exposure. Explicitly including water molecules in a simulation is a natural and rigorous but computationally expensive way to account for solvent effects. Implicit solvent models use approximations to gain computational efficiency but suffer from less accuracy. When comparable results can be achieved, implicit solvent models are more attractive because of their high computational efficiency. For the 37-residue FBP28 WW domain, a 20-ns run on 4 Intel Pentium 4/2.4 GHz processors required 1.5 days with the NPSA implicit solvent model compared with 1 month with an explicit solvent model.

The solvent accessible surface area-dependent term in the NPSA model is designed to account implicitly for solvation effects. The solvation parameter σ can tune the extent of atomic exposure to solvent by being more negative for more exposure and more positive for more burial. Backbone nitrogen atoms are more exposed in a β-extended structure than in an α-helical structure. This means that setting a more negative solvation parameter σ to backbone nitrogen atoms might lead to stabilization of β-extended structures. We thus modified the solvation parameter σ_{N_bone} of backbone nitrogen atoms from the original −0.06 kcal/mol·Å^2 to −0.12 kcal/mol·Å^2. We first conducted 20-ns simulations at 300 K for three runs with different heating speeds. For comparison, we also conducted simulations with an explicit water model at 300 K. The three-stranded β-sheet native structure was maintained very well with both the explicit water model and the modified NPSA model. The only difference was in the C-terminal residues Gln34, Glu35, and Leu36. These three residues formed a stable non-native 3−10 helix in all three trajectories with the explicit water model, whereas they sampled both extended and helical conformations with the modified NPSA model. The ϕ/φ distributions of the three residues in the three trajectories of each model are shown in Figure 8. We can see that the three residues mainly sampled helical regions with the explicit solvent model, whereas, with the modified NPSA model, the three residues mainly sampled

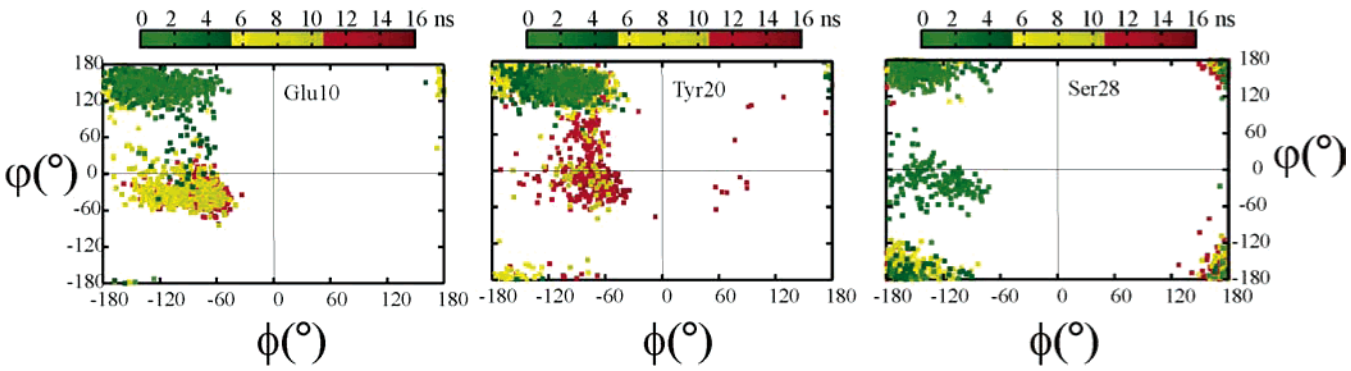


Figure 9. ϕ/φ distributions of the Glu10, Tyr20, and Ser28 residues of FBP28 in the simulation with $\sigma_{N\_bone} = -0.12$ kcal/mol·Å^2 in the NPSA model at 430 K. Glu10, Tyr20, and Ser28 are the middle residues of the first, the second, and the third strands in the native structure, respectively. The colors show the simulation time: beginning (green) to end (red).

the diagonal extended regions. In the NMR ensemble of 10 structures, the three C-terminal residues are either extended or coiled.

To investigate the performance of the modified NPSA model in unfolding simulations, we conducted 20-ns simulations at 430 K for three runs with different heating speeds. The three-stranded β-sheet native structure was lost in all three runs, and the remaining structural content was diverse with both non-native strands and helices existing. However, the loss of the native structure in the three runs did not exhibit an unfolding pathway consistent with experimental data. Figure 9 shows the time development of the ϕ/φ distributions of the middle residues Glu10, Tyr20, and Ser28 in the three native strands in one of the three 20-ns trajectories. The colors show the simulation time: beginning (green) to end (red). We can see that both Glu10 and Tyr20 moved to helical regions by the end of the simulation and Glu10 was faster than Tyr20 whereas Ser28 moved to corner extended regions. This reveals that the modified solvation parameter $\sigma_{N\_bone} = -0.12$ kcal/mol·Å^2 is not sufficient to prevent conversion to helical conformations and, at the same time, that the relative stabilities of the three strands were changed and became inconsistent with experimental data.

The lack of the expected improvement with the modified solvation parameter aiming at stabilizing extended conformations by increasing the solvent exposure of backbone nitrogen atoms indicates again that the intrinsic helix preference problem in AMBER force fields cannot be easily solved by refining a single parameter. In addition, we found that the modification of the solvation parameter in the NPSA model did not yield notable changes to the simulations for the helical trp-cage protein at both room temperature and higher temperatures. This indicates that the AMBER ff03 force field with the NPSA model is more robust for helical proteins than β-sheet proteins.

Conclusions

We have investigated several generations of the AMBER force field for simulation of a native 3-stranded β-sheet protein domain (along with several other small protein structures). The widely used ff94 force field and two backbone torsion potential variants were found to destabilize the β-sheet structure by directly converting it to an α-helix at high temperatures with the NPSA implicit solvent model. In contrast, the ff96 force field resulted in a highly elevated unfolding temperature and substantial non-native β-sheet structures with the NPSA implicit solvent model. These results are in agreement with previous studies revealing the helix-preference in the ff94 force field and the β-sheet preference in the ff96 force field with both an explicit solvent model and the generalized Born implicit solvent model. The newer ff03 force field showed much lower α-helix propensity compared to ff94 and lower β-strand propensity compared to ff96. More importantly, the ff03 force field allowed the observation of the experimentally observed unfolding pathway of the three-stranded β-sheet protein FBP28 with both the NPSA implicit solvent model and explicit water model. However, the ff03 force field still favors helical conformations in unfolded states. A modification of the solvation parameter of backbone nitrogen atoms in the NPSA model did not improve the results. This investigation also suggests that one should consider the integrative effects of all the force field parameters to improve the secondary structure balance of a force field.

Acknowledgment. The authors thank the Klaus Tschira Foundation for financial support and Dr. Peter J. Winn for critical reading of the paper.

Supporting Information Available: The full list of NPSA partial atomic charges. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J.; Kollman, P. A second generation of force field for the simulation of proteins, nucleic acids and organic molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.

(2) Kollman, P.; Dixon, R. W.; Cornell, W. D.; Fox, T.; Chipot, C.; Pohorille, A. The development/application of a 'minimalist' organic/biochemical molecular mechanic force field using a combination of ab initio calculations and experimental data. *Computer simulations of biological systems*; Escom: The Netherlands, 1997; pp 83−96.

(3) Wang, J.; Cieplak, P.; Kollman, P. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **2000**, *21*, 1049−1074.

(4) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **2003**, *24*, 1999−2012.

(5) Neria, E.; Fischer, S.; Karplus, M. Simulation of activation free energies in molecular systems. *J. Chem. Phys.* **1996**, *105*, 1902−1921.

(6) 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. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.

(7) van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Huenenberger, P. H.; Krueger, P.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G. *Biomolecular Simulation: The GROMOS Manual and User Guide*: Zuerich, 1996.

(8) Jorgensen, W. L.; Tirado-Rives, J. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* **1988**, *110*, 1657−1666.

(9) Simmerling, C.; Strockbine, B.; Roitberg, A. E. All-atom structure prediction and folding simulations of a stable protein. *J. Am. Chem. Soc.* **2002**, *124*, 11258−11259.

(10) Garcia, A. E.; Sanbonmatsu, K. Y. α-helical stabilization by side chain shielding of backbone hydrogen bonds. *Proc. Natl. Acad. Sci.* **2002**, *99*, 2782−2787.

(11) Sorin, E. J.; Pande, V. S. Empirical force-field assessment: The interplay between backbone torsions and noncovalent term scaling. *J. Comput. Chem.* **2005**, *26*, 682−690.

(12) Beachy, M.; Chasman, D.; Murphy, R.; Halgren, T.; Friesner, R. 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.

(13) Gnanakaran, S.; Garcia, A. E. Helix-coil transition of alanine peptides in water: force field dependence on the folded and unfolded structures. *Proteins* **2005**, *59*, 773−782.

(14) Zhou, R.; Berne, B. J. Can a continuum solvent model reproduce the free energy landscape of a beta -hairpin folding in water? *Proc. Natl. Acad. Sci.* **2002**, *99*, 12777−12782.

(15) Okur, A.; Strockbine, B.; Hornak, V.; Simmerling, C. Using PC clusters to evaluate the transferability of molecular mechanics force fields for proteins. *J. Comput. Chem.* **2003**, *24*, 21−31.

(16) Sorin, E. J.; Pande, V. S. Exploring the helix-coil transition via all-atom equilibrium ensemble simulations. *Biophys. J.* **2005**, *88*, 2472−2493.

(17) Wang, T.; Wade, R. C. Implicit solvent models for flexible protein−protein docking by molecular dynamics simulation. *Proteins* **2003**, *50*, 158−169.

(18) Ferguson, N.; Johnson, C. M.; Macias, M.; Oschkinat, H.; Fersht, A. Ultrafast folding of WW domains without structured aromatic clusters in the denatured state. *Proc. Natl. Acad. Sci.* **2001**, *98*, 13002−13007.

(19) Crane, J. C.; Koepf, E. K.; Kelly, J. W.; Gruebele, M. Mapping the transition state of the WW domain beta-sheet. *J. Mol. Biol.* **2000**, *298*, 283−292.

(20) Jaeger, M.; Nguyen, H.; Crane, J. C.; Kelly, J. W.; Gruebele, M. The folding mechanism of a beta-sheet: The WW domain. *J. Mol. Biol.* **2001**, *311*, 373−393.

(21) Ferguson, N.; Pires, J. R.; Toepert, F.; Johnson, C. M.; Pan, Y.-P.; Volkmer-Engert, R.; Schneider-Mergener, J.; Daggett, V.; Oschkinat, H.; Fersht, A. Using flexible loop mimetics to extend phi-value analysis to secondary structure interactions. *Proc. Natl. Acad. Sci.* **2001**, *98*, 13008−13013.

(22) Nguyen, H.; Jager, M.; Moretto, A.; Gruebele, M.; Kelly, J. W. Tuning the free-energy landscape of a WW domain by temperature, mutation, and truncation. *Proc. Natl. Acad. Sci.* **2003**, *100*, 3948−3953.

(23) Nymeyer, H.; Garcia, A. E. Simulation of the folding equilibrium of alpha-helical peptides: a comparison of the generalized Born approximation with explicit solvent. *Proc. Natl. Acad. Sci.* **2003**, *100*, 13934−13939.

(24) Ono, S.; Nakajima, N.; Higo, J.; Nakamura, H. Peptide free-energy profile is strongly dependent on the force field: Comparison of C96 and AMBER95. *J. Comput. Chem.* **2002**, *21*, 748−762.

(25) Frishman, D.; Argos, P. Knowledge-based protein secondary structure assignment. *Proteins* **1995**, *23*, 566−579.

(26) Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33−38, 27−38.

(27) Munoz, V.; Serrano, L. Elucidating the folding problem of helical peptides using empirical parameters. *Nat. Struct. Biol.* **1994**, *1*, 399−409.

(28) Rost, B. PHD: predicting one-dimensional protein structure by profile based on neural networks. *Methods Enzymol.* **1996**, *266*, 525−539.

(29) Neidigh, J. W.; Fesinmeyer, R. M.; Andersen, N. H. Designing a 20-residue protein. *Nat. Struct. Biol.* **2002**, *9*, 425−430.

CT0501607

JCTC Journal of Chemical Theory and Computation

Predicting Lattice Energy of Organic Crystals by Density Functional Theory with Empirically Corrected Dispersion Energy

Shaoxin Feng and Tonglei Li*

Pharmaceutical Sciences, University of Kentucky, Lexington, Kentucky 40536

Received August 1, 2005

Abstract: Calculation of the lattice energy of organic crystals is needed for predicting important structural and physicochemical properties such as polymorphism and growth morphology. Quantum mechanical methods that can be used for calculating typical organic crystals are unable to fully estimate van der Waals energies in a crystal. A method by augmenting the density functional theory with an analytical, nonelectronic approach for accounting for the dispersion energy was tested for selected organic crystals. The results illustrate the feasibility of this method for the prediction of the lattice energy of organic crystals. It is also shown that the dispersion energy is a dominant component of the lattice energy, particularly for those organic crystals that have no hydrogen bonds.

Introduction

Organic or molecular crystals play a central role in the pharmaceutical and fine chemical industry. Their structures and particulate properties greatly affect the handling and processing of materials and considerably control the performance of final products.[1] Because of the relatively weak intermolecular interactions, organic crystals are susceptible to the formation of polymorphs due to a change or disruption in the crystal growth environment.[2] Solvents, additives, impurities, supersaturation, and temperature are among key factors that affect how organic molecules pack in the solid state. Crystal packing of the same molecules may vary dramatically, resulting in different physical and chemical properties. Lattice energies of different polymorphs, however, can have similar values, making the prediction a very challenging task.[3] Consequently, calculation of the lattice energy not only offers a possible way for polymorph prediction but may also help understand the supramolecular chemistry and self-assembly during the nucleation and crystal growth processes.

A few methods can be used for calculating the lattice energy of molecular crystals. One often used is molecular mechanics based upon empirical force fields, which are constituted by a set of analytical equations using the positions and types of atoms as well as their bonding information for estimating interatomic interactions with the help of empirical parameters. Many have been developed for small molecules and biomolecules, such as Amber[4] and Dreiding.[5] Several potential models have been developed for periodic systems.[6] Nonetheless, being a totally empirical method, a force field has inherited difficulties for providing reliable energy estimations, especially for those structures that vary greatly from those used to develop the force field. Because the lattice energy of molecular crystals is relatively small, in particular, the difference between polymorphs can be as small as 2 kJ/mol,[2] or even smaller, which is beyond the typical accuracy of force-field based methods, calculating the lattice energy with force fields alone may pose a significant challenge for the polymorph prediction.

Quantum mechanical methods, on the other hand, may be capable of producing highly accurate energy estimations for a molecular system. However, one of the biggest challenges for calculating organic crystals in practice stems from the difficulty of fully considering the long-range van der Waals (vdW) energy.[7−9] As a quantum-mechanical phenomenon, vdW energies indicate mutually induced or correlated motions of electrons by the Coulomb interactions between atoms, even when the atoms are distantly apart.[7] The Hartree−Fock (HF) theory considers no such correlation

* Corresponding author phone: (859)257-1472; fax: (859)257-7585; e-mail: tonglei@uky.edu. Corresponding author address: 514 College of Pharmacy, University of Kentucky, 725 Rose Street, Lexington, KY 40536-0082.

10.1021/ct050189a CCC: $33.50
Published on Web 12/06/2005

energies; the density functional theory (DFT),[10−12] in principle, gives the exact description of ground-state energy, including the vdW energy. However, practical implementations relying on estimation strategies for the exchange-correlation functionals, including local density approximation (LDA)[11] and generalized gradient approximation (GGA),[13−15] cannot satisfyingly predict the vdW energies.[8] Using localized electron densities or their gradients fails to reproduce the physics of vdW interactions at large separations between atoms where there is little or no overlap of their electron densities. Higher-level quantum mechanical theories, such as MP2 (second-order Møller−Plesset perturbation theory[16]), are able to do a better job in considering the vdW energies, but they are very computationally demanding, making their applications for organic crystals impractical (except for a few simple crystal systems, such as C_2H_2 and CH_3OH[17,18]).

There have been many efforts for improving the quantum mechanical methods to account for the long-range vdW energies, including the introduction of vdW functionals to the traditional DFT methods.[7,19] One interesting approach among the efforts for the practical calculation of intermolecular interactions is to augment the HF and DFT methods with analytical models of vdW potentials parametrized empirically, in a similar way as those being used in molecular mechanics.[20] The augmentation, not part of the electronic calculations and only based on positions and types of nuclei, accommodates the quantum mechanical methods posteriorly through the adjustment of the empirical vdW models. It was applied to the HF[21−23] and recently to the DFT.[9] It appears to be a practical and flexible approach for considering the vdW energies at large interatomic distances but to damp or tune down at small distances where the HF or DFT takes over and can carry out reliable calculations of intermolecular energies.

Because the London dispersion force is a major, universal contributor to the vdW force,[24] dispersion energy is often equally quoted as the long-range vdW energy. It is argued, however, that the vdW interactions also include the Keesom force (due to the orientation effect between permanent dipoles) and the Debye force (due to the induction effect between a permanent dipole and an induced dipole),[24] which may be trivial as compared to the dispersion force. For our purpose to study the lattice energies of organic crystals, we will use the two concepts, dispersion energy and long-range (attractive) vdW energy, exchangeably.[25] As the lattice energy of an organic crystal consists of short-range, electrostatic, induction (polarization), and dispersion energies,[26] the dispersion energy is believed to be significant, especially for crystals with no hydrogen bonds present. In this study, we present calculation results of lattice energies of selected organic crystals by using the empirically augmented DFT method and discuss possible future improvements.

Methodology

Lattice energy, E_{latt}, of an organic crystal is the energy difference between the bulk crystal, E_{xtal}, and isolated molecule, E_{mol}, of the same compound:

$$E_{latt} = E_{xtal} - E_{mol} \quad (1)$$

It is the energy requirement for vaporizing a crystal, representing the cohesive or intermolecular interactions in the solid state. Negative values of lattice energy indicate attractive intermolecular interactions of a crystal. Conversely, positive values indicate repulsive interactions. The calculation of lattice energy of selected organic crystals was carried out in two separated steps. Nondispersive energies were calculated by DFT first, followed by estimation of the dispersion energy with an empirical method.

For selected organic crystals, their crystal structures were obtained from Cambridge Structural Database.[27] Single-point energy calculations were conducted with DFT after the structural optimization where lattice parameters were kept the same as experimental values, while the system energy was minimized with respect to the fractional coordinates of atoms. DFT with B3LYP exchange-correlation functional[13,28] was used for the structural optimization and energy calculation. When calculating the energy, the basis set superposition error (BSSE)[29] was considered by the counterpoise method.[30] Fifty ghost atoms were typically placed around each atom within 5 Å in order to obtain acceptable BSSE corrections. Furthermore, the single molecule of each compound was optimized independently with the DFT-B3LYP method so that the possible energy reduction due to conformational change of the molecule from the solid state to the gas phase could be considered. The effect of basis sets was studied as well. A periodic ab initio program, Crystal 03,[31] was used for the optimization and single-point electronic calculations. The energy convergence of the structural optimizations and single-point electronic calculations was set as 10^{-7} Hartree. The root-mean-squares (RMS) of energy gradient and atomic displacement were set to 0.0003 and 0.0012 atomic units, respectively. All calculations were performed on a 16-CPU Linux cluster.

The dispersion energy between a pair of atoms at long range can be evaluated by a power series of the interatomic distance, R[26]

$$E_{disp}(R) = -\sum_{n=6}^{\infty} C_n R^{-n} \quad (2)$$

where n are even numbers, and C_n are dispersion coefficients. The first term, C_6R^{-6}, is the dominant contribution, representing the instantaneous dipole-instantaneous dipole interaction,[8] and is often used in practice as the only term of dispersion energy. The subsequent terms (C_8R^{-8}, $C_{10}R^{-10}$, etc.) are attributed to interactions between higher-order fluctuating multipole moments.

To preserve the true nature of dispersion energy that is not infinite at $R = 0$, a damping function is often used to correct the power series in eq 2 for calculating the dispersion energy.[26] A general form of the damping functions remains one at long range and decays to zero when $R = 0$. Various types of damping functions have been reported.[9,21,23,32,33] In this study, the form of dispersion energy is give by

$$E_{disp}(R) = -f_d(R)C_6R^{-6} \quad (3)$$

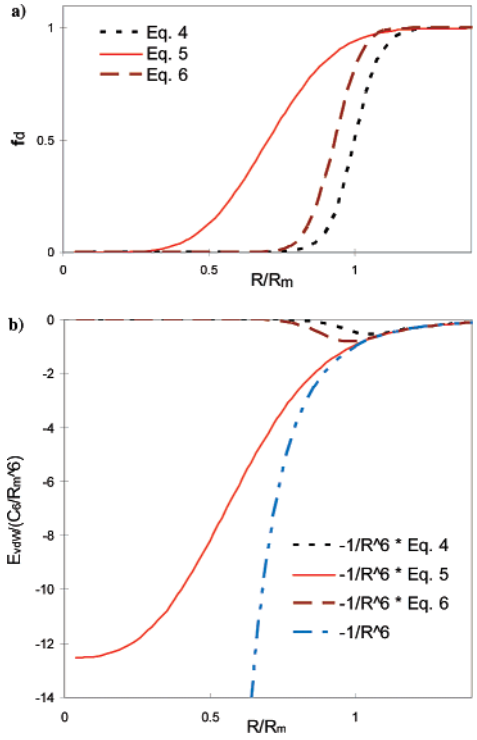


Figure 1. Plots of the three damping functions (a) and their influences on the van der Waals interaction (b).

where the damping function, $f_d(R)$, may take the following forms[9]

$$f_d(R) = \frac{1}{1 + \exp\left[-D_1\left(\frac{R}{R_m} - 1\right)\right]} \quad (4)$$

$$f_d(R) = \left(1 - \exp\left[-D_2\left(\frac{R}{R_m}\right)^3\right]\right)^2 \quad (5)$$

where R_m is the damping radius, taken as the sum of atomic van der Waals radii[34] of the pair of atoms. The coefficients, D_1 and D_2, are associated with the quality of the damping functions; they were assigned to 23.0 and 3.54, respectively, by Wu and Yang.[9] Eq 5 was used by Mooij et al as well but with D_2 as 7.19.[35] Other formats of damping functions include one by Elstner et al[33]

$$f_d(R) = \left(1 - \exp\left[-D_3\left(\frac{R}{R_m}\right)^7\right]\right)^4 \quad (6)$$

where D_3 was assigned as 3.0. These three damping functions are shown in Figure 1, along with their effects on a general C_6R^{-6} term. It appears that the damping strength by eq 6 is between those by eqs 4 and 5 with eq 4 being the strongest. All these functions were tested in this study.

Accurate values of intermolecular C_6 coefficients can be obtained experimentally from the dipole oscillator strength distributions[36,37] or computationally from the frequency-dependent polarizabilities.[38,39] It is not a trivial task, however, to decompose the intermolecular C_6 coefficients into interatomic C_6 coefficients. Most often, interatomic C_6 coefficients are produced by data fitting, either directly to intermolecular C_6 coefficients[9] or indirectly to molecular polarizabilities.[20,40] In this study, the interatomic C_6 coefficients reported by Wu and Yang were used without modification in our calculations of dispersion energies. In their method,[9] interatomic C_6 coefficients were obtained by least-squares fitting to intermolecular C_6 coefficients that were accurately determined experimentally.[36,37] It was assumed that an intermolecular coefficient was the additive sum of interatomic coefficients of all atom pairs of the molecules. It was also assumed that different molecules could use the same set of atomic C_6 coefficients. Their tests of evaluating molecular pairs indicated that the interatomic C_6 coefficients developed in their study were able to produce satisfying results for augmenting the DFT method. Nonetheless, it is argued that these empirically derived interatomic dispersion coefficients, generalized over atom types, may be limited and inflexible in dealing with other systems in which the molecular environment is greatly varied from the training sets.[41]

In this study, dispersion energies of bulk crystals were evaluated atom−atom pairwisely with eq 3. The cutoff distance for considering an atom pair was set to 25 Å, which only had a difference of 0.01 kJ/mol or less when compared to energy values calculated without any cutoff. Urea and benzene were used for evaluating the effects of basis sets, BSSE, and conformational change of molecules between the crystal and gas phase. Thirty-three organic crystals were selected for testing the method. Experimentally determined sublimation enthalpies of these compounds available in the literature were used for providing experimental estimates of calculated lattice energies. Sublimation enthalpy, $\Delta H_{sub}(T)$, and lattice energy, E_{latt}, are related to each other by the following equation

$$\Delta H_{sub}(T) = -E_{latt} - E_0 + \int_0^T \Delta C_p dT \quad (7)$$

where T is the temperature at which the sublimation enthalpy is measured, E_0 is the zero-point energy, and ΔC_p is the difference in heat capacity between the gas and solid phases. As the sublimation enthalpy of a crystal is typically determined from its vapor pressures at different temperatures, using the above equation to derive the lattice energy is practically difficult. ΔC_p is temperature-dependent; the heat capacity of solids is very small at low temperature, requiring more than routine instrumental methods. The equation implies that the lattice energy is a quantity determined at the absolute zero, disregarding the thermal contributions. By ignoring the zero-point energy (normally less than 1% of E_{latt}[42]) as well as making assumptions that the gas phase is ideal and contributions from intramolecular vibrations are equal in the solid and gas phases, an approximation form to derive lattice energy from sublimation enthalpy is given by[43]

$$\Delta H_{sub}(T) = -E_{latt} - 2RT \quad (8)$$

where R is the gas constant. In this study, experimental values of lattice energies were estimated by the above equation from literature data of sublimation enthalpies.

Results and Discussion

Using the DFT-B3LYP/6-31G** method, the lattice energy of urea was calculated as −95.14 kJ/mol. The conformational

change alone of urea from the crystal to gas phase accounted for 20.09 kJ/mol. The symmetry, C_{2v}, of urea in the solid state was reduced to C_2 during the optimization of the single molecule, the most stable conformer in the gas phase.[44] If the symmetry was kept, the energy contribution of the conformational change was 12.88 kJ/mol instead. Therefore, full optimization of single molecules in the gas phase seems to be essential and was carried out for other crystals as well in this report. The BSSE that was associated with DFT-B3LYP/6-31G** was estimated with 94 ghost atoms as 32.00 kJ/mol. After considering the BSSE, the lattice energy of urea calculated by DFT was −63.14 kJ/mol. Furthermore, based on the optimized urea crystal structure, dispersion energies were calculated by eq 3 with the three damping functions defined by eqs 4−6 as −39.03, −57.09, and −43.30 kJ/mol, respectively. Added together from values calculated by DFT and each of the three analytical methods, lattice energies were −102.17, −120.23, and −106.44 kJ/mol. The experimental value of sublimation enthalpy of urea was reported as 98.6 kJ/mol,[45] leading to its lattice energy estimated by eq 8 as −103.6 kJ/mol. The calculation results of the lattice energies appear to be acceptable, particularly using the dispersion energies calculated by the damping functions, eqs 4 and 6. More importantly, it is clearly indicated that the lattice energy calculated by DFT alone is greatly underestimated. In the case of urea, the dispersion energy accounts for about 40% of the total lattice energy.

The effect of basis sets was studied with urea and benzene. Lattice energies of urea calculated by DFT-B3LYP with 6-21G, 6-21G**, 6-31G, 6-31G**, 6-311G, and 6-311G** after BSSE corrections of 86.72, 88.41, 34.92, 32.00, 27.64, and 28.69 kJ/mol were −79.50, −54.96, −80.88, −63.14, −81.24, and −59.54 kJ/mol, respectively. The absolute values of nondispersive energies were significantly smaller by the polarized basis sets (6-21G**, 6-31G**, and 6-311G**) than those by the nonpolarized basis sets (6-21G, 6-31G, and 6-311G). This resulted in the fact that the polarized basis sets had the total energy of the single molecule decreased more than that of the crystal. The less sensitive effect by the Gaussian type basis sets on the energies of periodic systems stems from the "real" basis sets used in the calculation being Bloch functions, which have the periodicity of the crystal lattice and have their "local" functions built up with linear combinations of the Gaussian type basis sets. For the same reason, extended basis sets with diffuse orbitals can cause numerical instabilities and are suggested not to be used. Moreover, lattice energies of benzene calculated with 6-21G, 6-21G**, 6-31G, 6-31G**, 6-311G, and 6-311G** after BSSE corrections of 27.33, 28.26, 18.15, 18.93, 9.02, and 8.45 kJ/mol were 13.78, 13.83, 16.26, 15.47, 17.43, and 16.41 kJ/mol, respectively. Positive values imply repulsive intermolecular interactions (without the consideration of dispersion energy). Polarized basis sets gave similar energy values and trend lines as those by nonpolarized basis sets, likely due to the nonpolar feature of the benzene molecule. It is interesting to note that the HF/6-21G** and HF/6-31G** gave smaller BSSE values of urea and benzene than ours.[46] Balancing the accuracy and computing time, therefore, the 6-21G** basis set was used for the calculations of other organic crystals.

Table 1 lists the results of lattice energies of 33 organic crystals, including nondispersive energies calculated by DFT-B3LYP/6-21G** with the BSSE correction, dispersion energies by the three damping functions, and experimental values of sublimation enthalpy and derived lattice energy data. The reference codes of the crystals and the temperatures under which crystals structures were determined are listed in Table 2. Dispersion energies calculated with the damping function, eq 6, produced absolute values between larger ones by eq 5 and smaller ones by eq 4 and gave the closest lattice energies to experimentally derived data. From the percentage in the lattice energies, the dispersion energies are a major component of intermolecular interactions of the organic crystals. Cyanamide in Table 1 has the smallest value, 42%, which is already a significant number. Most crystals that have hydrogen bonds have the dispersion energy between 40 and 65% of their lattice energies. The percentage is significantly higher for crystals that have no hydrogen bonds between their molecules in crystal. For those crystals whose percentages of dispersion energy are more than 100%, their nondispersive energies (E_{DFT} + BSSE) are positive, meaning that the conformation of individual molecules in the crystals is likely to be energy-unfavorable due to close contacts. It can be further noticed that the crystals that have positive or very absolutely small nondispersive energies have no hydrogen bonds. The integrity of the crystals is likely to be kept solely by the dispersion energy. It should be noted that the optimization of crystal structures was carried out without the consideration of dispersion energy; it was done purely based on nondispersive energies. It is believed that the nondispersive energies calculated by DFT in an organic crystal are responsible for the conformation of individual molecules, while the dispersion energy plays a key role in deciding the volume of unit cell, especially for a crystal that has no hydrogen bonding. In fact, compared to the system energy of a crystal calculated by DFT, the dispersion energy is trivial (e.g., the dispersion energy is about 3×10^{-5} of the total DFT energy of urea). It is very likely that the introduction of dispersion energy during the structural optimization of a crystal may have little influence on the fractional coordinates of atoms, but affect the lattice constants. Consequently, the lattice constants of a crystal were kept the same as the experimental values during optimizations. As shown in Table 2, the root-mean-square (RMS) values due to the optimization of atomic Cartesian coordinates of all crystals studied are small, indicating that the optimization method is sound and the exclusion of dispersion energy is acceptable. The major contribution to the RMS values appears to be a result of position changes of H atoms. This is not surprising since most X-ray diffraction measurements are not able to directly determine fractional coordinates of H atoms. Thus, it is thought that during the optimization, especially when the lattice parameters are kept constant and the space group is maintained, the close contacts or short-range interactions between atoms, not the long-range, collective van der Waals energy, play a more important role in determining the fractional coordinates of atoms. Still, the

Table 1. Calculated Energy Values of Selected Organic Crystals, Including Nondispersive (E_{DFT}), BSSE, Dispersion (E_{disp}), and Lattice Energies (E_{latt})[m]

	E_{DFT}	BSSE	E_{DFT} + BSSE	E_{disp} (eq 4)	E_{disp} (eq 5)	E_{disp} (eq 6)	E_{latt} (eq 6)	ΔH_{sub}	E_{latt} (eq 8)
acetamide*	−120.90	80.76	−40.14	−40.26	−55.47	−44.66	−84.80	77.2[a]	−82.2
anthracene	−13.41	49.37	35.96	−124.18	−137.65	−133.44	−97.48	103.4[b]	−108.4
benzene	−14.43	28.26	13.83	−58.10	−65.88	−63.63	−49.80	44.4[c]	−49.4
1,2-benzene-dicarbonitrile	−61.19	57.09	−4.10	−71.42	−81.56	−77.28	−81.38	86.9[d]	−91.9
benzoic acid*	−91.54	81.58	−9.96	−75.11	−94.67	−82.14	−92.10	89.7[b]	−94.7
1,1′-biphenylene	−11.91	37.46	25.55	−99.31	−109.66	−107.07	−81.52	87.3[e]	−92.3
chrysene	−21.12	51.57	30.45	−142.72	−154.25	−152.21	−121.76	118.8[f]	−125.2
cyanamide*	−92.96	44.64	−48.32	−31.14	−44.17	−34.80	−83.12	75.2[a]	−80.2
cyanoacetamide*	−134.22	77.31	−56.91	−51.77	−68.45	−57.65	−114.56	100.4[a]	−105.4
cyanuric acid*	−173.77	114.35	−59.42	−64.11	−92.07	−72.24	−131.66	133.6[a]	−138.6
cyclohexane	3.51	22.97	26.48	−69.76	−75.00	−73.94	−47.46	46.6[g]	−49.7
1,4-cyclohexanedione	−100.78	100.46	−0.32	−67.61	−83.67	−75.47	−75.79	84.2[a]	−89.2
dicyanodiamide*	−154.66	75.28	−79.38	−56.77	−77.43	−62.71	−142.09	129.3[a]	−134.3
diglycolid anhydride	−105.44	86.08	−19.36	−53.10	−65.49	−59.49	−78.85	84.0[a]	−89.0
1,3-dinitrobenzene	−93.80	97.33	3.53	−79.93	−94.23	−88.07	−84.54	81.2[h]	−86.2
formamide*	−116.11	73.99	−42.12	−29.84	−44.52	−33.82	−75.94	71.7[a]	−76.7
furan 2,5-dicarboxylic acid*	−193.45	132.16	−61.29	−78.19	−104.11	−85.57	−146.86	125.7[a]	−130.7
imidazole*	−85.31	45.70	−39.61	−48.74	−61.27	−53.35	−92.96	80.8[a]	−85.8
maleic anhydride	−74.79	66.17	−8.62	−45.65	−55.30	−51.47	−60.09	68.1[a]	−73.1
naphthalene	−9.93	40.46	30.53	−93.81	−105.01	−102.10	−71.57	72.6[b]	−77.6
propanoic acid*	−93.03	69.96	−23.07	−44.36	−56.87	−47.88	−70.95	74.0[i]	−77.8
pyrazine	−54.78	55.12	0.34	−53.64	−62.16	−57.77	−57.43	56.2[j]	−61.2
pyrazole*	−70.62	41.89	−28.73	−48.04	−62.69	−54.10	−82.83	71.7[a]	−76.7
squaric acid*	−198.30	117.19	−81.11	−55.49	−84.42	−64.20	−145.31	154.3[a]	−159.3
succinic acid*	−190.66	143.71	−46.95	−69.98	−94.98	−77.68	−124.63	123.1[a]	−128.1
succinic anhydride	−101.02	89.29	−11.73	−51.99	−68.04	−59.74	−71.47	82.3[a]	−87.3
tetracyanomethane	−62.32	56.69	−5.63	−42.76	−68.16	−52.73	−58.36	61.1[k]	−66.1
1,3,5-triazine	−55.40	54.73	−0.67	−48.39	−54.00	−52.26	−52.93	56.7[a]	−61.7
2,4,5-trimethylbenzoic acid*	−83.38	69.69	−13.69	−92.41	−108.21	−97.75	−111.44	109.6[l]	−114.6
1,3,5-trioxane	−100.73	99.22	−1.51	−49.96	−58.76	−55.07	−56.58	55.6[a]	−60.6
urea	−95.14	32.00	−63.14	−39.03	−57.09	−43.30	−106.44	98.6[a]	−103.6
urethane*	−108.71	77.40	−31.31	−49.82	−63.20	−53.72	−85.03	76.3[a]	−81.3
urotropine	−75.43	86.95	11.52	−94.37	−101.64	−100.46	−88.94	79.0[a]	−84.0

[a] Reference 45. [b] Reference 48. [c] Reference 49. [d] Reference 50. [e] Reference 51. [f] Reference 52, T = 383 K. [g] Reference 53, T = 186K. [h] Reference 54. [i] Reference 55, T = 225−238 K. [j] Reference 56, T = 288−317 K. [k] Reference 57. [l] Reference 58. [m] E_{DFT} and BSSE were calculated with DFT-B3LYP/6-21G** except for urea which was calculated with 6-31G**. Crystals that have hydrogen bonds are marked with asterisks. Sublimation enthalpies (ΔH_{sub}) and derived lattice energies are also listed. Unless indicated otherwise, the sublimation enthalpies were measured at 298 K. Energy unit: kJ/mol.

full optimization by considering the dispersion energy is necessary for correcting the temperature effect on the lattice volume, since most X-ray structural determinations are typically carried out under ambient conditions or in the range of 100−200 K (Table 2).

The results of lattice energy are also plotted in Figure 2 along with experimental values. The correlation coefficient of the calculated and experimental data, r^2, is 0.92, when eq 6 was used for calculating the dispersion energy. The coefficient became 0.79 or 0.87 if eq 4 or eq 5 was used, respectively. As shown in Figure 1, the damping function of eq 6 is not as quick as eq 4 to tune down the van der Waals interaction and is not as slow as eq 5 either when the interatomic distance decreases. The damping strength of such a function appears to be a key factor in controlling the quality of the calculation of dispersion energies. It can also be seen from Figure 2 that both crystals with and without hydrogen bonds have similar matching qualities to experimental values of the lattice energy. Crystals with hydrogen bonds may have better calculated lattice energies, except for three crystals with the largest calculated values, furan 2,5-dicarboxylic acid, squaric acid, and dicyanodiamide, which also have the largest nondispersive energies. In addition, the majority of crystals without hydrogen bonds have their calculated lattice energies absolutely smaller than the experimental values, suggesting that the damping function, eq 6, may underestimate the dispersion energy. Consequently, the better match of crystals with hydrogen bonds to their experimental values implies that the DFT method (B3LYP/6-21G**) may overestimate the nondispersive energy, canceling out the error of dispersion energy by the damping function. Since the lattice energy accounts for the intermolecular interactions in a crystal (eq 1), the overestimation by DFT is likely due to the BSSE which may not reach the convergence because of ghost atoms being insufficient. This clearly needs to be considered in the future studies. Considering the fact that the DFT method used in this study may not be the best method and there is always a better one that can generally produce more accurate nondispersive energies, if a better DFT method is used for calculating the nondispersive energy, the damping function

Table 2. Reference Codes of the Calculated Crystals in the Cambridge Structural Database and Temperatures under Which the Crystal Structures Were Determined[a]

	ref code	temp (K)	RMS (Å) (excluding H)	RMS (Å) (H only)	RMS (Å)
acetamide*	ACEMID05	23	0.142	0.195	0.174
anthracene	ANTCEN09	94	0.044	0.134	0.093
benzene	BENZEN01	138	0.197	0.373	0.298
1,2-benzene-dicarbonitrile	YUYPUD01	153	0.197	0.305	0.233
benzoic acid*	BENZAC07	20	0.134	0.169	0.149
1,1′-biphenylene	BIPHNE01	130	0.057	0.138	0.098
chrysene	CRYSEN	283−303	0.032	0.130	0.086
cyanamide*	CYANAM01	108	0.143	0.163	0.151
cyanoacetamide*	CYANAC	283−303	0.190	0.251	0.216
cyanuric acid*	CYURAC05	100	0.036	0.032	0.035
cyclohexane	CYCHEX	115	0.069	0.218	0.182
1,4-cyclohexanedione	CYHEXO	133	0.159	0.244	0.206
dicyanodiamide*	CYAMPD03	83	0.045	0.079	0.061
diglycolid anhydride	DLGYAH	283−303	0.174	0.206	0.185
1,3-dinitrobenzene	DNBENZ11	100	0.141	0.222	0.165
formamide*	FORMAM02	90	0.142	0.179	0.162
furan 2,5-dicarboxylic acid*	FURDCA	283−303	0.274	0.337	0.292
imidazole*	IMAZOL06	103	0.133	0.145	0.138
maleic anhydride	MLEICA01	130	0.163	0.269	0.191
naphthalene	NAPHTA15	100	0.086	0.142	0.114
propanoic acid*	PRONAC	178	0.168	0.344	0.278
pyrazine	PYRAZI01	184	0.035	0.208	0.134
pyrazole*	PYRZOL05	108	0.036	0.178	0.122
squaric acid*	KECYBU06	283−303	0.022	0.026	0.023
succinic acid*	SUCACB09	130	0.071	0.193	0.137
succinic anhydride	SUCANH12	100	0.265	0.251	0.260
tetracyanomethane	TCYETY11	283−303	0.283	N/A	0.283
1,3,5-triazine	TRIZIN02	283−303	0.024	0.066	0.043
2,4,5-trimethylbenzoic acid*	RUVQAA	283−303	0.116	0.200	0.164
1,3,5-trioxane	TROXAN11	103	0.039	0.128	0.095
urea	UREAXX02	148	0.025	0.046	0.037
urethane*	ECARBM01	168	0.136	0.287	0.230
urotropine	HXMTAM10	15	0.012	0.008	0.010

[a] Root-mean-square (RMS) values of atomic Cartesian coordinates of each crystal due to the structural optimization are also listed. DFT-B3LYP/6-21G** was used for the optimization.

needs to be tailored with regard to diminishing the van der Waals potential than those used in this study. More suitable interatomic C_6 coefficients are also needed to be developed from various means. A recent report by Johnson and Becke suggested a general model for developing C_6 coefficients without the empirical fitting.[41] Because of the empirical nature of calculating the dispersion energy as well as the lack of knowledge of the "true" value of the dispersion energy of an organic crystal, the coupling between the quantum mechanical and empirical methods for predicting the lattice energy will remain challenging, requiring significant experimental inputs.

Unfortunately, there are unavoidable experimental errors and systematic variances that are associated not only with the determination of sublimation enthalpy but also with the derivation of lattice energy. It is common to see disagreement between sublimation enthalpies of the same materials in the literature.[47] The discrepancy can be caused by different instrumentations, different research groups, and even different ways to prepare the materials. Defects and impurities can greatly affect the thermodynamic properties. To directly determine the sublimation enthalpy by measuring the vapor pressure of the solid at different temperatures may run into troubles of possible solid−solid phase transitions as well as difficulties to accurately detect the (extremely low) vapor pressure. To indirectly estimate the sublimation enthalpy by using a thermodynamic cycle and measuring the fusion enthalpy and vaporization enthalpy can be challenging due to the lack of sufficient data on heat capacity as well as the uncertainties associated with correcting the data.[47] For example, the sublimation enthalpy of anthracene at 298 K has been reported many times ranging from 85 to 105 kJ/mol, while the recommended value is 103.4 kJ/mol (Table 1).[48] Furthermore, using eq 8 adds uncertainties to the estimation of lattice energies. It is estimated that the contribution by heat capacity to the sublimation enthalpy is no more than 10% of lattice energy.[42] Thus, given the fact that the heat capacities of most organic solids are not available at low temperature, using the correction of *2RT* may contribute a systematic error no greater than 5% for crystals that have sublimation enthalpies ranged around 100 kJ/mol at 298 K.

Clearly, the prediction of lattice energies of organic crystals requires advances in both computational and ex-

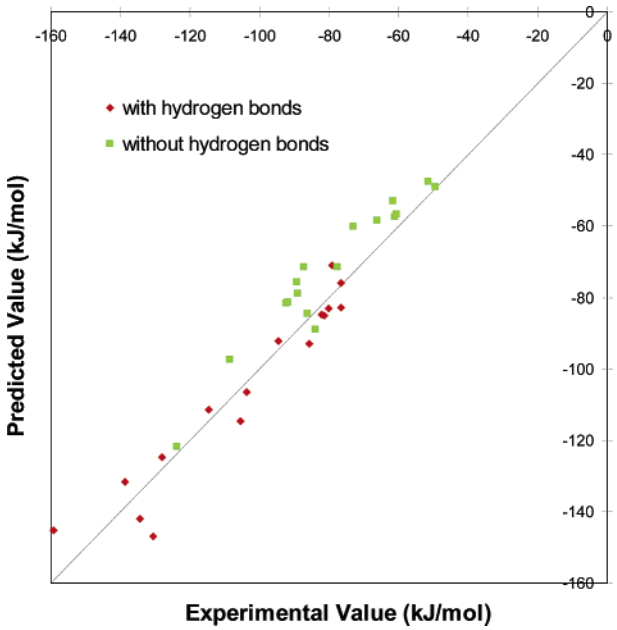


Figure 2. Comparison of calculated and experimental lattice energies. Organic crystals with and without hydrogen bonds are denoted differently. The dispersion energy was evaluated with eq 6.

perimental methods. The surprisingly good agreement between the calculated and experimental values of lattice energy (Figure 2) indicates that the prediction method of lattice energy appears to be sound and reliable. Due to the empirical nature, the damping function and the C_6 coefficients for estimating the dispersion energy should be adjusted in accordance to the DFT method that is used for calculating the nondispersive energy and, ideally, should be tailor-made for each crystal system. A recent report sheds some light in the improvement of the empirical model (eq 2).[41]

Conclusions

Lattice energies of selected organic crystals were calculated by DFT with subsequent corrections by empirically calculated dispersion energies. The calculated results show a good agreement with experimentally estimated values. By taking the C_6R^{-6} dispersion energy into account at large interatomic distance but diminishing it at small distance, the empirical method is likely to compensate the inability of routine DFT methods for fully describing electron correlations at the region where electron clouds are not closely overlapped. Choosing a proper analytical damping function as well as the interatomic C_6 coefficients is important for producing high-quality data. Clearly, due to the empirical nature of posterior corrections, the methodology requires both computational and experimental improvements for the prediction of lattice energies of organic crystals.

Acknowledgment. This work was supported by NSF (DMR-0449633).

References

(1) Byrn, S. R.; Pfeiffer, R. R.; Stowell, J. G. *Solid-State Chemistry of Drugs*, 2nd ed.; SSCI, Inc.: West Lafayette, IN, 1999.

(2) Hollingsworth, M. D. *Science* **2002**, *295*, 2410−2413.

(3) Price, S. L. *Adv. Drug Deliv. Rev.* **2004**, *56*, 301−319.

(4) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; Debolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. *Comput. Phys. Commun.* **1995**, *91*, 1−41.

(5) Mayo, S. L.; Olafson, B. D.; Goddard, W. A. *J. Phys. Chem.* **1990**, *94*, 8897−8909.

(6) Gale, J. D.; Rohl, A. L. *Mol. Simul.* **2003**, *29*, 291−341.

(7) Kohn, W.; Meir, Y.; Makarov, D. E. *Phys. Rev. Lett.* **1998**, *80*, 4153−4156.

(8) Dobson, J. F.; McLennan, K.; Rubio, A.; Wang, J.; Gould, T.; Le, H. M.; Dinte, B. P. *Aust. J. Chem.* **2001**, *54*, 513−527.

(9) Wu, Q.; Yang, W. T. *J. Chem. Phys.* **2002**, *116*, 515−524.

(10) Hohenberg, P.; Kohn, W. *Phys. Rev. B* **1964**, *136*, B864-&.

(11) Kohn, W.; Sham, L. J. *Phys. Rev.* **1965**, *140*, 1133-&.

(12) Kohn, W.; Becke, A. D.; Parr, R. G. *J. Phys. Chem.* **1996**, *100*, 12974−12980.

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

(14) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648−5652.

(15) Perdew, J. P.; Burke, K.; Wang, Y. *Phy. Rev. B* **1996**, *54*, 16533−16539.

(16) Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618−622.

(17) Nagayoshi, K.; Ikeda, T.; Kitaura, K.; Nagase, S. *J. Theor. Comput. Chem.* **2003**, *2*, 233−244.

(18) Nagayoshi, K.; Kitaura, K.; Koseki, S.; Re, S. Y.; Kobayashi, K.; Choe, Y. K.; Nagase, S. *Chem. Phys. Lett.* **2003**, *369*, 597−604.

(19) Dion, M.; Rydberg, H.; Schroder, E.; Langreth, D. C.; Lundqvist, B. I. *Phys. Rev. Lett.* **2004**, *92*.

(20) Halgren, T. A. *J. Am. Chem. Soc.* **1992**, *114*, 7827−7843.

(21) Ahlrichs, R.; Penco, R.; Scoles, G. *Chem. Phys.* **1977**, *19*, 119−130.

(22) Aziz, R. A.; Chen, H. H. *J. Chem. Phys.* **1977**, *67*, 5719−5726.

(23) Hepburn, J.; Scoles, G.; Penco, R. *Chem. Phys. Lett.* **1975**, *36*, 451−456.

(24) French, R. H. *J. Am. Ceram. Soc.* **2000**, *83*, 2117−2146.

(25) Dobson, J. F.; Wang, J.; Dinte, B. P.; McLennan, K.; Le, H. M. *Int. J. Quantum Chem.* **2005**, *101*, 579−598.

(26) Buckingham, A. D.; Fowler, P. W.; Hutson, J. M. *Chem. Rev.* **1988**, *88*, 963−988.

(27) Allen, F. H. *Acta Crystallogr., Sect. B: Struct. Sci.* **2002**, *58*, 380−388.

(28) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098−3100.

(29) Davidson, E. R.; Feller, D. *Chem. Rev* **1986**, *86*, 681−696.

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

(31) Dovesi, R.; Orlando, R.; Civalleri, B.; Roetti, C.; Saunders: V. R.; Zicovich-Wilson, C. M. *Z. Kristallogr.* **2005**, *220*, 571−573.

(32) Tang, K. T.; Toennies, J. P. *J. Chem. Phys.* **1984**, *80*, 3726−3741.

(33) Elstner, M.; Hobza, P.; Frauenheim, T.; Suhai, S.; Kaxiras, E. *J. Chem. Phys.* **2001**, *114*, 5149−5155.

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

(35) Mooij, W. T. M.; van Duijneveldt, F. B.; van Duijneveldt-van de Rijdt, J. G. C. M.; van Eijck, B. P. *J. Phys. Chem. A* **1999**, *103*, 9872−9882.

(36) Kumar, A.; Meath, W. J. *Chem. Phys.* **1994**, *189*, 467−477.

(37) Kumar, A.; Meath, W. J. *Mol. Phys.* **1997**, *90*, 389−398.

(38) Spackman, M. A. *J. Chem. Phys.* **1991**, *94*, 1295−1305.

(39) Stanton, J. F. *Phys. Rev. A* **1994**, *49*, 1698−1703.

(40) Miller, K. J. *J. Am. Chem. Soc.* **1990**, *112*, 8533−8542.

(41) Johnson, E. R.; Becke, A. D. *J. Chem. Phys.* **2005**, *123*.

(42) Gavezzotti, A. In *Structure Correlation*; Burgi, H.-B., Dunitz, J. D., Eds.; VCH: Weinheim, 1994; Vol. 2, pp 509−542.

(43) Gavezzotti, A.; Filippini, G. In *The Molecular Solid State: Theoretical Aspects and Computer Modeling*; Gavezzotti, A., Ed.; John Wiley & Sons: New York, 1997; pp 61−98.

(44) Masunov, A.; Dannenberg, J. J. *J. Phys. Chem. A* **1999**, *103*, 178−184.

(45) Dewit, H. G. M.; Vanmiltenburg, J. C.; Dekruif, C. G. *J. Chem. Thermodyn.* **1983**, *15*, 651−663.

(46) Spackman, M. A.; Mitchell, A. S. *Phys. Chem. Chem. Phys.* **2001**, *3*, 1518−1523.

(47) Chickos, J. S.; Acree, W. E. *J. Phys. Chem. Ref. Data* **2002**, *31*, 537−698.

(48) Sabbah, R.; An, X. W.; Chickos, J. S.; Leitao, M. L. P.; Roux, M. V.; Torres, L. A. *Thermochim. Acta* **1999**, *331*, 93−204.

(49) Dekruif, C. G. *J. Chem. Thermodyn.* **1980**, *12*, 243−248.

(50) Satotoshima, T.; Sakiyama, M.; Seki, S. *Bull. Chem. Soc. Jpn.* **1980**, *53*, 2762−2767.

(51) Osborn, A. G.; Scott, D. W. *J. Chem. Thermodyn.* **1980**, *12*, 429−438.

(52) Nass, K.; Lenoir, D.; Kettrup, A. *Angew. Chem. Int. Ed. Engl.* **1995**, *34*, 1735−1736.

(53) Bondi, A. *J. Chem. Eng. Data* **1963**, *8*, 371−381.

(54) Jones, A. H. *J. Chem. Eng. Data* **1960**, *5*, 196−200.

(55) Calisvanginkel, C. H. D.; Calis, G. H. M.; Timmermans, C. W. M.; Dekruif, C. G.; Oonk, H. A. J. *J. Chem. Thermodyn.* **1978**, *10*, 1083−1088.

(56) Sakoguchi, A.; Ueoka, R.; Kato, Y.; Arai, Y. *Kagaku Kogaku Ronbunshu* **1995**, *21*, 219−223.

(57) Barnes, D. S.; Mortimer, C. T.; Mayer, E. *J. Chem. Thermodyn.* **1973**, *5*, 481−483.

(58) Colomina, M.; Jimenez, P.; Perezossorio, R.; Roux, M. V.; Turrion, C. *J. Chem. Thermodyn.* **1987**, *19*, 155−162.

CT050189A

JCTC Journal of Chemical Theory and Computation

Molecular Dynamics Simulation of Iminosugar Inhibitor−Glycosidase Complex: Insight into the Binding Mechanism of 1-Deoxynojirimycin and Isofagomine toward β-Glucosidase

Jin-Ming Zhou,*,† Jun-Hong Zhou,‡ Yi Meng,† and Min-Bo Chen*,†

Department of Computer Chemistry and Cheminformatics, Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences, 354 Fenglin Lu, 200032, Shanghai, China, and Chemical and Pharmaceutical Institute, East China University of Science and Technology, 130 Meilong Lu, 200237, Shanghai, China

Received July 11, 2005

Abstract: The binding mechanism of iminosugar inhibitor 1-deoxynojirimycin and isofagomine toward β-glucosidase was studied with nanosecond time scale molecular dynamics. Four different systems were analyzed according to the different protonated states of inhibitor and enzyme (acid/base carboxyl group, Glu166). The simulations gained quite a reasonable result according to the thermodynamic experimental fact. Further conclusions were made including the following: (1) 1-deoxynojirimycin binds with the β-glucosidase as conjugate acid forms; (2) the slow onset inhibition of isofagomine aims to slow deprotonation of the acid/base carboxyl group which is caused by a nearly zero hydrogen bond interaction between the hydroxyls of the acid/base carboxyl group; and (3) the nucleophile carboxyl group plays an important role when the inhibitor binds with glucosidase.

Introduction

The inhibitors of glycosidases are subject to intense current interest for they not only serve as important tools for studying the biological functions of oligosaccharides and the hydrolysis mechanism of glycosidases but also are prospective therapeutic agents for a variety of carbohydrate-mediated diseases.[1−3] The iminosugar achieved by the ring oxygen or anomeric carbon of pyranose or furanose replaced by the imino group is a kind of most potent glycosidase inhibitor.[4−8] These unique molecules promise a new generation of iminosugar-based medicines in a wide range of diseases such as diabetes,[9] viral infections,[10] tumor metastasis,[11] and lysosomal storage disorders.[12]

Of them, 1-deoxynojirimycin **(1, Chart 1)** and isofagomine **(2,** Chart 1**)** are of particular interest in inhibitor design for they and their derivatives are often powerful inhibitors of glycosidase action.[5,6,13,14] The mechanism of inhibition is thought of as their conjugate acid mirrors positive charge development at the endocyclic oxygen or the anomeric carbon of the glycosidase transition state **(3,** Chart 1**)** so that they can gain a tight binding complex with the glycosidase enzyme.[15] Though there is little structure difference between **1** and **2**, the inhibiting properties of them toward β-glycosidase are rather dissimilar: (1) **2** is a much more stronger inhibitor than **1** for sweet almond β-glycosidase and some other β-glycosidases.[16] (2) A slow-onset inhibition is observed for **2** when binding with the enzyme, while **1** is as a linear steady-state rate.[17,18] (3) Recent van't Hoff analysis of temperature dependence of binding of **1** and racemic **2** to sweet almond β-glycosidase shows that the binding of **1** toward the enzyme was enthalpically driven, while the binding of **2** with the enzyme was with unfavorable enthalpy and was actually entropically driven.[16] This adds the odds to the inhibition mechanism that the affinity of an enzyme for a transition-state mimic should necessarily be driven by

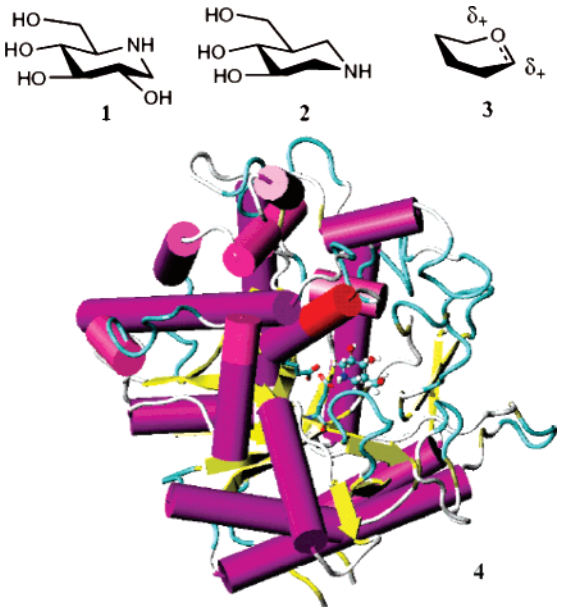
* Corresponding author phone: 86-021-54925277; fax: 86-021-64166128; e-mail: zhoujm@mail.sioc.ac.cn (J.-M.Z.) or mbchen@mail.sioc.ac.cn (M.-B.C.).

† Chinese Academy of Sciences.

‡ East China University of Science and Technology.

10.1021/ct050168g CCC: $33.50
Published on Web 11/09/2005

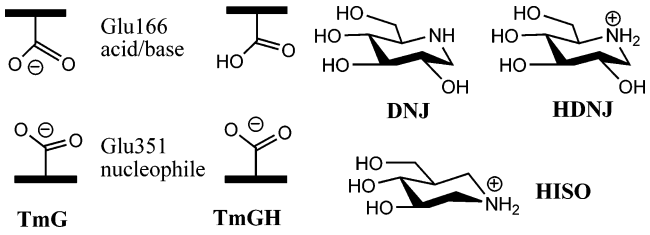
Chart 1. Transition-State Analogues 1-Deoxynojirimycin (**1**), Isofagomine (**2**), and Considerable Oxocarbenium Ion Transition State (**3**) and Crystal Structure of the Inhibitor and TmGH1β-Glucosidase Complex (**4**): Protein (Cartoon), Inhibitor (CPK), Glu166 and Glu351 (Bonds), Which Was Generated by VMD Soft Package



a large and favorable change in enthalpy, a criterion which was proposed by Wolfenden.[19] However, as a further study of the binding of **1** and **2** to β-glycosidase by David L. Zechel et al., in which the value of binding enthalpy was measured by the ITC (Isothermal Titration Calorimetry) method, similar favorable binding enthalpy changes were gained in contrast to the result by van't Hoff analysis, and it seems the large favorable entropy makes **2** a much better inhibitor of β-glycosidase than **1**. Otherwise, the crystal structures show that **1** binds with the TmGH1β-glycosidase in a skew-boatlike conformation while **2** is in a chair conformation, which indicates they may have different binding modes with β-glycosidase. It was suggested by the author that the superior inhibition of **2** relative to **1** is not the result of superior transition-state mimicry but benefits from an entropic advantage and a more favorable electrostatic interaction with the acid/base catalyst.[18]

The studies referred above, assuredly, gave close insight into the mechanism of inhibition. However, there are still a few questions remaining intangible. What causes an entropic advantage of **2** comparing to **1** when binding with β-glycosidase? An explanation was given that such entropic advantage may be caused by the binding of **1** incorporating approximately 1−3 more water molecules at the molecular interface relative to the binding of **2**. This has been observed in the crystal structure, but no evidence was shown in solution. **2** show a slow-onset inhibition toward the enzyme while **1** does not. The explanation of a slow conformational change in the enzyme or an unusual change of the ionization state of the inhibitor residues has been proposed, and each has some evidences. However, do they have some relationship, which means that the conformational change in the enzyme may perhaps be caused by the change of ionization

Chart 2. Label of the Glucosidase and Inhibitors: Glucosidase with Glu166 Protonated (TmGH), Glucosidase with Glu166 Unprotonated (TmG), 1-Deoxynojirimycin (DNJ), the Conjugate Acid of 1-Deoxynojirimycin (HDNJ), and the Conjugate Acid of Isofagomine (HISO)



state of the inhibitor residues? Molecular dynamics in the water box can always get some useful information of the dynamic property of protein in water, and it is also an effective way to study the folding and unfolding or changes of conformation of protein in water. [20,21] Therefore, molecular dynamics will be a particularly suitable means to explore the problems which have been mentioned above.

In particular, **1** (pK_a, 6.7) would be largely unprotonated when entering the active site under pH 6−7, and it has long been argued that **1** may bind to glucosidases as a neutral amine rather than a protonated conjugate acid.[18,22,23] Additionally, as for the glucosidase, the pK_a values of the carboxyl groups of acid/base (Glu166) and nucleophile (Glu351) deduced from the pH dependence of K_{cat}/K_m are respectively 6.96 and 4.75,[18] which indicates that the nucleophile carboxyl group would be mainly unprotonated, while the acid/base carboxyl group could be both unprotonated and protonated species. So the mostly possible combined mode would be TmGH-DNJ or TmG-HDNJ. Otherwise, isofagomine (pK_a, 8.6) would be expected to be largely protonated when entering the active site at pH 6−7. Evidence from the crystal structure of the complex shows that the iminosugar is protonated within the active site, and the two carboxyl groups of acid/base and nucleophile are both unprotonated.[24] Therefore, the process of inhibitor binding with glucosidase should include deprotonation of the acid/base carboxyl group (Glu166). So in our study, the simulations TmGH-DNJ, TmG-HDNJ, TmGH-HISO, and TmG-HISO (label shown as Chart 2), summarized in Table 1, were performed.

Material and Methods

The models of TmG-HDNJ and TmGH-DNJ were built up based on the X-ray crystal of complex of **1** and TmGH1β-glycosidase at 2.2 Å resolution (PDB entry code, loim). [18] The missing residues (Ser1, Asn2, Glu233) and many other missing atoms were repaired according to the X-ray crystal structure of TmGH1β-glycosidase (PDB entry code 1od0)[18] with the molecular modeling software package Sybyl 6.9 (Tripos Inc.). The hydrogens of HDNJ and DNJ were added using the build/edit menu which is included in the Sybyl software package, and then the charge of the structures were calculated at the B3LYP/6-311++G** level using the "pop=CHelpG" keywords[25] in the Gaussion98a software package. [26] The models of TmG-HISO and TmGH-HISO were built up based on the X-ray crystal of complex of **2**

Table 1. Summary of the Molecular Dynamics Simulations

simulation label	solute (glucosidase and inhibitor)	number of water	simulation length (ns)
TmGH-DNJ	TmGH1 (Glu166 protonated), 1-deoxynojirimycin	16159	6
TmG-HDNJ	TmGH1 (Glu166 unprotonated), conjugate acid of 1-deoxynojirimycin	16159	6
TmGH-HISO	TmGH1 (Glu166 protonated), conjugate acid of Isofagomine	16788	6
TmG-HISO	TmGH1 (Glu166 unprotonated), conjugate acid of Isofagomine	16788	6

and TmGH1 β-glycosidase at 2.2 Å resolution (PDB entry code, 1oif).[18] The missing residues (Ser1, Asn2) and atoms repair of protein and the hydrogen addition of HISO as well as the charge calculation were done by the same way as models TmG-HDNJ and TmGH-DNJ. Topology files were generated using the pdb2gmx program included in the GROMACS software package and OPLS force field parameters were applied except for the charge upon the inhibitors (HDNJ, DNJ, and HISO). [27] The hydrogen atoms of the protein were also added. For TmG-HDNJ and TmG-HISO, the residue Glu166 was set unprotonated. For TmGH-DNJ and TmGH-HISO, Glu166 was set protonated. Each model was solvated with SPC water molecules in a cube box and ensured the whole surface of the protein to be covered by a water layer with a thickness more than 12 Å. Several (9 or 10) Na^+ ions were added to the system to keep it zero net charge.

The energy minimization for each model was performed using the steepest descent algorithm (100 steps), followed by the conjugate gradient (1000 steps) in the GROMACS 3.1.4 software package.[28] Then a 100 ps position restrained molecular dynamics was performed with the protein and inhibitor fixed in order to let the waters and Na^+ equilibrate around them. Finally, a 6-ns molecular dynamics was started by taking initial velocities from a Maxwellian distribution at 300 K. Solvent and solute were independently, weakly coupled to a temperature bath with a relaxation time of 0.1 ps. The system was also isotropically, weakly coupled to a pressure bath at 1.0 atm with a relaxation time of 0.5 ps and an isothermal compressibility of 0.45×10^{-4}. [29] Long-range electrostatics was calculated with the particle-mesh Ewald method. [30] Short-range van der Waals and Coulombic interactions were cut off at 1.0 and 1.0 nm, respectively. All bond lengths were constrained using the LINCS algorithm,[31] and the time step was set to 0.002 ps. When the molecular dynamics were finished, analyses were performed using facilities within the GROMACS package.

The binding free energy between the inhibitor and the enzyme was calculated using the LIE (linear interaction energy) method developed by Åqvist et al.[32,33] The LIE method is based on the assumption that, using MD or Monte Carlo conformational simulations, the binding free energy of an inhibitor to a receptor target can be expressed as the equation

$$\Delta G_{\text{bind}} = \alpha[\langle V_{\text{i-s}}^{\text{vdw}}\rangle_{\text{bound}} - \langle V_{\text{i-s}}^{\text{vdw}}\rangle_{\text{free}}] + \beta[\langle V_{\text{i-s}}^{\text{ele}}\rangle_{\text{bound}} - \langle V_{\text{i-s}}^{\text{ele}}\rangle_{\text{free}}] + \gamma$$

where $\langle\ \rangle$ denotes MD or MC averages of the nonbonded van der Waals (vdw) and electrostatic (ele) interactions between the inhibitor and its surrounding environment (i-s), i.e., either the solvated receptor binding site (bound state) or just the solvent (free state). α and β are the scaling factors for the averaged van der Waals energies and the averaged electrostatic energies. The scaling factors α β tend to be system dependent, and γ is always set to zero. For our calculation, α is set to 0.181 for all systems, and β is set to 0.33 or 0.50 when the inhibitor is a neutral amine or a protonated conjugate acid, respectively, based on Åqvist et al.'s work.[34]

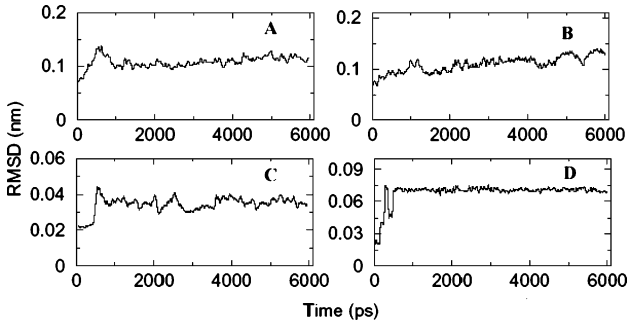


Figure 1. Time dependence of Cα RMSD of the protein with respect to the crystal structure: (A) simulation TmG-HDNJ and (B) simulation TmGH-DNJ. Time dependence of RMSD of the ligand with respect to the crystal structure: (C) simulation TmG-HDNJ and (D) simulation TmGH-DNJ. All curves are obtained by 30 ps average.

Results

The Simulations of TmG-HDNJ and TmGH-DNJ. As described above, TmGH-DNJ or TmG-HDNJ may be the most possible combined mode for **1** binding with the enzyme. Simulations TmGH-DNJ and TmG-HDNJ were performed to verify which binding mode would be mostly likely. The root-mean-square deviation of between the instantaneous MD and crystal structure was reported in Figure 1. Both simulations reach a structural equilibrium after about 1200 ps, and the RMSD values of Cα of protein are not beyond 0.15 nm, which indicates that the protein structure in solution has a small deviation from that in the crystal. As for the simulation TmG-HDNJ, the RMSD values of Cα of protein keep quite stable after it reaches a structural equilibrium while the RMSD values of HDNJ keep fluctuating around 0.04 nm, and the ring conformation of inhibitor keeps a skew-boatlike form. For simulation TmGH-DNJ, the RMSD values of Cα of protein fluctuate in a range of about 0.05 nm after 4 ns, while the RMSD values of DNJ keep fluctuating around 0.06 nm, and the ring conformation of the inhibitor stays in a chairlike form. Thus, the RMSD analysis indicates some differences between simulation TmG-HDNJ and TmGH-DNJ both in the dynamic property of protein and conformation of ligand.

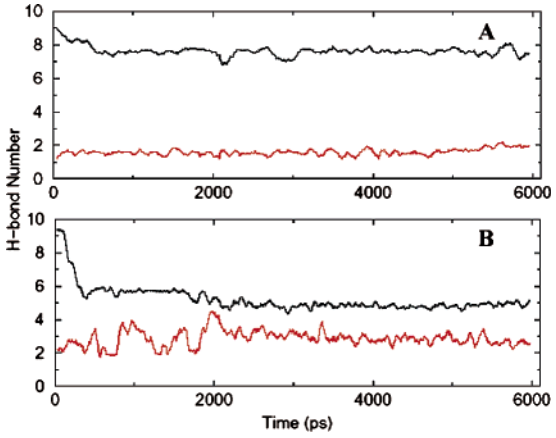


Figure 2. Time dependence of the number of hydrogen bonds: (A) simulation TmG-HDNJ, the number of hydrogen bonds between ligand and protein (black line), the number of hydrogen bonds between ligand and solvent (red line) and (B) simulation TmGH-DNJ, the number of hydrogen bonds between ligand and protein (black line), the number of hydrogen bonds between ligand and solvent (red line). All curves are obtained by 30 ps average.

The ligand has four hydroxyls and one amino (for TmGH-DNJ) or ammonium (for TmG-HDNJ) group, so the hydrogen bond interaction between the ligand and the protein as well as solvent molecules at the active site is very important for the inhibitor to bind tightly with the enzyme. Time dependence of the number of hydrogen bonds was shown in Figure 2. For simulation TmG-HDNJ, the number of hydrogen bonds between the ligand and the protein is around 8 after equilibrium, and the hydrogen bond number between the ligand and the solvent is around 2. However, for simulation TmGH-DNJ, the number of hydrogen bonds between the ligand and the protein is around 5 after equilibrium, and the number of hydrogen bonds between the ligand and the solvent is around 3. All together, the TmG-HDNJ binding mode has about two more hydrogen bonds than the TmGH-DNJ binding mode, which would cause more tightly binding between the ligand and receptor.

Snapshots of the active site structure of both simulations at 4 ns were shown in Figure 3, which can give more details about the difference of the two combined modes. As for simulation TmG-HDNJ, when the ligand is the conjugate acid of **1**, the residues Gln20, Asn165, Glu351, and Glu405 form hydrogen bonds with a ligand directly, while Glu166 forms a solvent-mediated hydrogen bond with the ligand. Of them, Glu351 is extremely important, for it forms a strong hydrogen bond with both 2-hydroxyl and ammonium of the conjugate acid. Otherwise, there are two water molecules forming a hydrogen bond with the ligand. The pyranoid ring of the ligand is distorted as a skew-boat conformation and fits the structure in crystal very well. As for simulation TmGH-DNJ, when the ligand is 1-deoxynojirimycin, His121, Glu166, Glu351, Glu405, and Trp406 form hydrogen bonds with the ligand directly and Asn165 forms a solvent-mediated hydrogen bond with the ligand, while Glu351 only forms a hydrogen bond with 2-hydroxyl and a solvent-mediated hydrogen bond with 3-hydroxyl of the ligand. There are also two hydrogen bonds between the ligand and the solvent

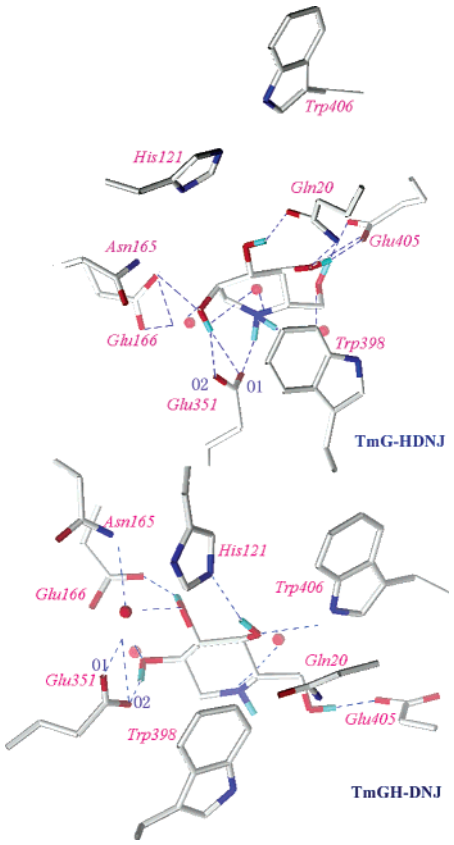


Figure 3. Snapshots of the structure of the active site of both simulations at 4 ns time step: up (simulation TmG-HDNJ) and down (simulation TmGH-HDNJ). Red balls are as water molecules.

molecule. The pyranoid ring conformation of the ligand of simulation TmGH-DNJ adopts a chairlike form rather than a skew-boat conformation for TmG-HDNJ, which does not fit that in the crystal structure.

For Glu351 plays an important role as the ligand binds with the receptor, the distances from amine or ammonium, 2-hydroxyl of the ligand to the two oxygen atoms of the carboxyl group of Glu351 were examined along all the MD simulation, and the result was reported in Figure 4. As for simulation TmG-HDNJ, the distance from O1 of the carboxyl group of Glu351 to H (N^+H_2) of the ligand is always shorter than 2.0 Å with the average of 1.6 Å; the distance from O1 of the carboxyl group of Glu351 to H (2-OH) of the ligand is between 2.0 and 3.0 Å with the average of 2.6 Å; the distance from O2 of the carboxyl group of Glu351 to H (2-OH) of the ligand fluctuates around 2.0 Å with the average of 1.9 Å. These indicate that there is a strong interaction between the residue Glu351 and the ligand and that such an interaction may contribute greatly to tightly binding of the ligand toward the receptor. As for simulation TmGH-DNJ, distances both from O1 of the carboxyl group of Glu351 to H (NH) of the ligand and from O2 of the carboxyl group of Glu351 to H (2-OH) of the ligand grow to more than 3.5 Å soon after the start of the simulation, due to the conformational change of the ligand from a skew-boatlike form to a chairlike form at about 200 ps; only the distance from O1 of the carboxyl group of Glu351 to H (2-OH) of the ligand stays below 2.0 Å and the average is 1.7 Å. Therefore,

Table 2. Average Interaction Energies (kJ/mol) between Ligand and Surroundings in the Bound and Unbound States in MD Simulation and Binding Free Energy Calculated by LIE (Linear Interaction Energy) Method Comparing with Observed Binding Free Energy Value

binding mode	$\langle V_{i-s}^{vdw}\rangle_{free}$	$\langle V_{i-s}^{vdw}\rangle_{bound}$	$\langle V_{i-s}^{el}\rangle_{free}$	$\langle V_{i-s}^{el}\rangle_{bound}$	ΔG_{bind}	observed ΔG_{bind}
TmGH-DNJ	-15.48 ± 1.22	-49.59 ± 3.38	-245.83 ± 2.81	-236.68 ± 0.36	-3.15 ± 2.42	-33.30 ± 0.00
TmG-HDNJ	-5.13 ± 0.12	-33.04 ± 1.79	-282.48 ± 0.42	-365.88 ± 4.43	-46.75 ± 2.77	

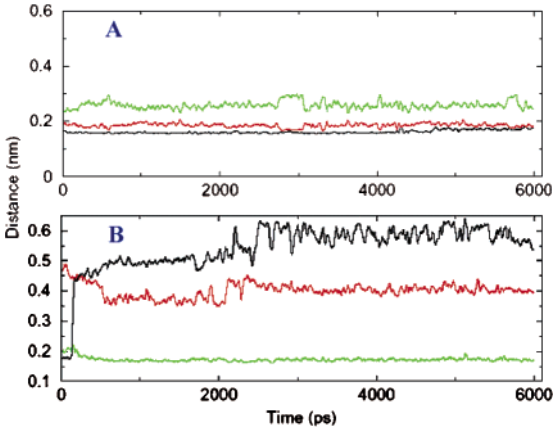


Figure 4. Time dependence of distances from amine or ammonium, 2-hydroxyl of the ligand to the two oxygen atoms of the carboxyl group of Glu351: (A) simulation TmG-HDNJ, distance from O1 of the carboxyl group of Glu351 to H (N^+H_2) of the ligand (black line), distance from O1 of the carboxyl group of Glu351 to H (2-OH) of the ligand (green line), distance from O2 of the carboxyl group of Glu351 to H (2-OH) of the ligand (red line) and (B) simulation TmGH-DNJ, distance from O1 of the carboxyl group of Glu351 to H (NH) of the ligand (black line), distance from O1 of the carboxyl group of Glu351 to H (2-OH) of the ligand (green line), distance from O2 of the carboxyl group of Glu351 to H (2-OH) of the ligand (red line). The atom label O1, O2 was shown as in Figure 3. All curves are obtained by 30 ps average.

Glu351 would contribute much less to the binding of the ligand toward the receptor when the ligand is the free amine rather than the conjugate acid.

The binding free energy was calculated using the LIE method for both binding modes TmG-HDNJ and TmGH-DNJ, respectively. Default scaling factors $\alpha = 0.181$ and $\beta = 0.50$ were used for TmG-HDNJ. For TmGH-DNJ, when the ligand is neutral, the value of β was set to 0.33. Shown in Table 2, the calculated binding free energies of the conjugate acid and the free amine are -46.75 and -3.15 kJ/mol, respectively. The observed binding energy is -33.30 kJ/mol,[18] which indicates that the conjugate acid should be the most possible state when 1-deoxynojirimycin binds with the β-glucosidase.

The Simulations of TmG-HISO and TmGH-HISO. The RMSD of both the protein and the ligand of simulation TmG-HISO and TmGH-HISO was reported in Figure 5. Also, both simulations reach a structural equilibrium after about 1200 ps. The RMSD values of Cα of the protein for both simulations stay below 0.15 nm. For simulation TmG-HISO, it is quite stable after it reaches a structural equilibrium. As for simulation TmGH-HISO, the RMSD values of Cα of the protein fluctuate between 0.1 and 0.15 nm after structural

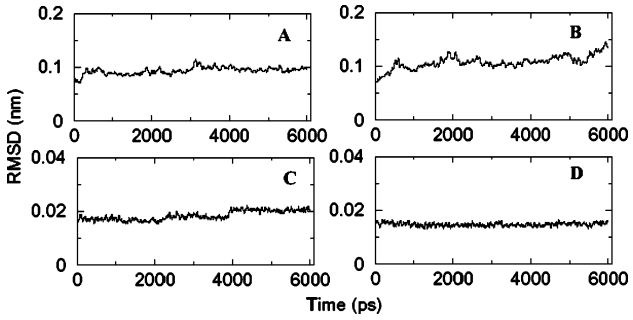


Figure 5. Time dependence of Cα RMSD of the protein with respect to the crystal structure: (A) simulation TmG-HISO and (B) simulation TmGH-HISO. Time dependence of RMSD of the ligand with respect to the crystal structure: (C) simulation TmG-HISO and (D) simulation TmGH-HISO. All curves are obtained by 30 ps average.

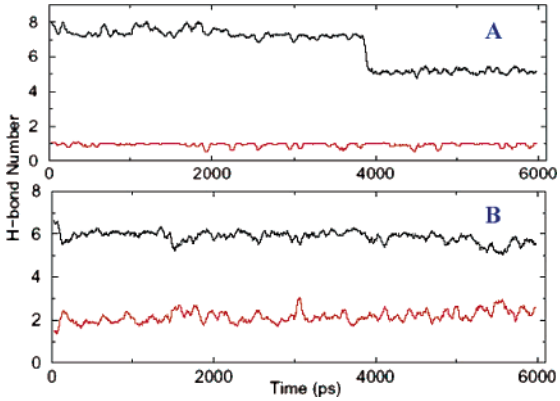


Figure 6. Time dependence of the number of hydrogen bonds: (A) simulation TmG-HISO, the number of hydrogen bonds between the ligand and the protein (black line), the number of hydrogen bonds between the ligand and the solvent (red line) and (B) simulation TmGH-HISO, the number of hydrogen bonds between the ligand and the protein (black line), the number of hydrogen bonds between the ligand and the solvent (red line). All curves are obtained by 30 ps average.

equilibrium. The RMSD values of the ligand of both simulations keep stable about 0.02 nm after equilibrium, consisting of the likewise chair $^1C_4$ conformations of the ligand in both simulations.

Time dependence of the number of hydrogen bonds was shown in Figure 6. For simulation TmG-HISO, the number of hydrogen bonds between the ligand and the protein is around 7 after equilibrium and then drops to 5 after about 4 ns, and the hydrogen bond number between the ligand and solvent is around 1. However, for simulation TmGH-HISO, the number of hydrogen bonds between the ligand and the protein is around 6 after equilibrium, and the number of

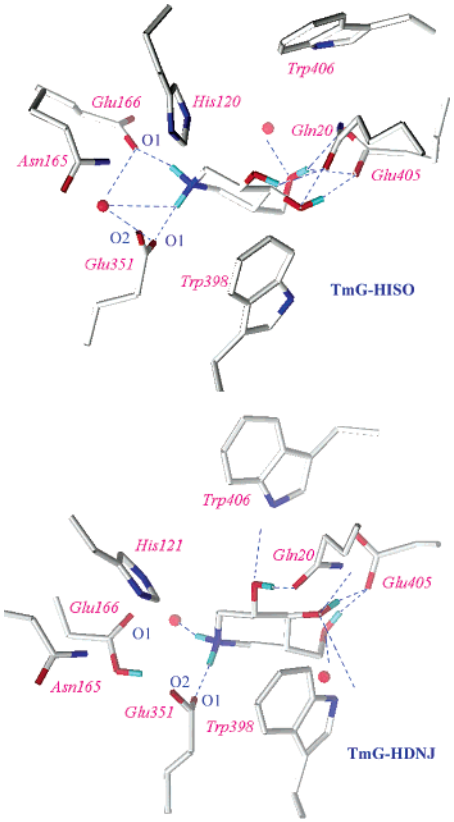


Figure 7. Snapshots of the structure of the active site of both simulations at 4 ns time step: up (simulation TmG-HISO) and down (simulation TmGH-HISO). Red balls are as water molecules.

hydrogen bonds between the ligand and the solvent is around 2. Snapshots of the active site structure of both simulations at 4 ns shown in Figure 7 provide more detailed information of the interaction between the ligand and the receptor. As for simulation TmG-HISO, when the residue Glu166 of the protein is unprotonated, the residues Gln20, Glu166, Glu351, and Glu405 form hydrogen bonds with the ligand directly, and both Glu166 and Glu351 form an additional water-mediated hydrogen bond with ligand. As for simulation TmGH-HISO, when the residue Glu166 of the protein is protonated, the residues Gln20, Glu351, Trp398, Glu405, and Glu406 form hydrogen bonds with the ligand directly, and no water-mediated hydrogen bond is found.

Residues Glu166 and Glu351 are very important for the ligand to bind tightly with the glucosidase for not only hydrogen bond interaction but also for strong electrostatic interaction between them. So the distances between the residue Glu351, Glu166, and the ligand were examined and reported in Figure 8. As for simulation TmG-HISO, distance from O1 of carboxyl group of Glu351 to H (N^+H_2) of ligand is always shorter than 2.0 Å and the average is 1.6 Å; distance from O2 of carboxyl group of Glu351 to H (N^+H_2) of ligand is always shorter than 2.0 Å and the average is 1.7 Å; distance from O1 of Glu166 to H (N^+H_2) of ligand fluctuates between 2.0 and 3.0 Å due to the rotation of torsion, then keep stable about 1.7 Å after around 2 ns. While, as for simulation TmGH-HISO, distance from O1 of carboxyl group of Glu351 to H (N^+H_2) of ligand is always shorter

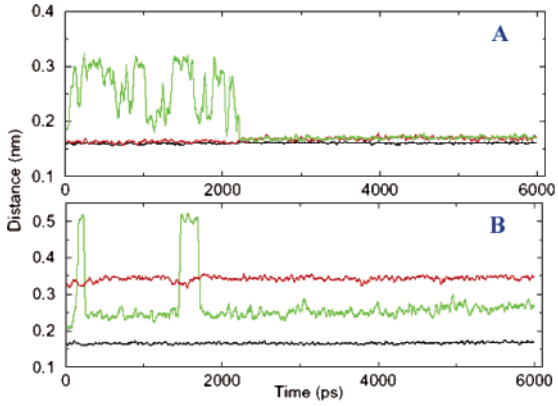


Figure 8. Time dependence of distances from ammonium of the ligand to the oxygen atoms of the carboxyl group of Glu351 and Glu166: (A) simulation TmG-HISO, distance from O1 of the carboxyl group of Glu351 to H (N^+H_2) of the ligand (black line), distance from O2 of the carboxyl group of Glu351 to H (N^+H_2) of the ligand (red line), distance from O1 of the carboxyl group of Glu166 to H (N^+H_2) of the ligand (green line) and (B) simulation TmGH-HISO, distance from O1 of carboxyl group of Glu351 to H (N^+H_2) of ligand (black line), distance from O2 of the carboxyl group of Glu351 to H (N^+H_2) of the ligand (red line), distance from O1 of the carboxyl group of Glu166 to H (N^+H_2) of the ligand (green line);

than 2.0 Å and the average is 1.7 Å; distance from O2 of carboxyl group of Glu351 to H (N^+H_2) of ligand fluctuates around 3.5 Å; distance from O1 of Glu166 to H (N^+H_2) of ligand fluctuates slightly around 2.5 Å except for two steep leap to 5.0 Å. From the result of distance analysis, we can see, when the Glu166 is unprotonated, the interaction between Glu166, Glu351, and the ligand would be much stronger than that when Glu166 is protonated, and this would compensate inferior position taken by the less number of hydrogen bond interaction. Results of binding free energy calculation were reported in Table 3. When the Glu166 of glucosidase is unprotonated, the binding free energy of ligand toward glucosidase is −41.87 kJ/mol, which is quite reasonable in contrast to the observed binding free energy − 45.60 kJ/mol. When the Glu166 of glucosidase is protonated, the binding free energy of ligand is −38.51 kJ/mol. As a comparison, the TmG-HISO binding mode is about 3.4 kJ/mol favored in contrast to the TmGH-HISO binding mode.

Discussion

The Binding Mode of 1-Deoxynojirimycin with β-Glucosidase. Though much of the published work assumes 1-deoxynojirimycin binds with glucosidase as the protonated form, which mimics the positive charge development at the anomeric carbons of the glycosidase transition state,[4,15] it has long been argued that 1-deoxynojirimycin may bind to glucosidases as a neutral amine rather than a protonated conjugate acid.[18,22,23] As the result of simulation TmG-HDNJ and TmGH-DNJ, we found that the MD snapshot structure fit better with the crystal structure when the ligand is a conjugate acid than when the ligand is a neutral amine. Furthermore, the binding free energy calculated by the LIE method indicates that the protonated ligand binds more tightly than the neutral ligand, and also a value of −46.76 kJ/mol

Table 3. Average Interaction Energies (kJ/mol) between Ligand and Surroundings in the Bound and Unbound States in MD Simulation and Binding Free Energy Calculated by LIE (Linear Interaction Energy) Method Comparing with Observed Binding Free energy value

binding mode	$\langle V_{l-s}^{vdw}\rangle_{free}$	$\langle V_{l-s}^{vdw}\rangle_{bound}$	$\langle V_{l-s}^{el}\rangle_{free}$	$\langle V_{l-s}^{el}\rangle_{bound}$	ΔG_{bind}	observed ΔG_{bind}
TmG-HISO	-6.77 ± 0.31	-38.14 ± 0.06	-251.83 ± 1.41	-324.21 ± 0.14	-41.86 ± 0.84	-45.60 ± 1.09
TmGH-HISO	-6.77 ± 0.31	-30.60 ± 0.45	-251.83 ± 1.41	-320.23 ± 0.81	-38.52 ± 1.25	

is more reasonable than that of the neutral ligand −3.15 kJ/mol in contrast to the observed binding free energy −33.30 kJ/mol. Both evidences indicate that 1-deoxynojirimycin may bind to glucosidases as protonated conjugate acid. A 6-ns molecular dynamic in water of 1-deoxynojirimycin was also performed, and the result shows that the equilibrium conformation of the ligand in water is in the chair form. Quantum calculation results also indicate that the chair conformation seems to be more stable than the boat conformation,[35] and the relative energy is about more than 20 kJ/mol. So in the bound state, when the interaction between the neutral amine of **1** and the protein is weak, the conformation of it would tend to be in the chair form, as a result of simulation TmGH-DNJ.

The Mechanism of Slow-Onset Inhibition of Isofagomine. It was proposed that slow-onset inhibition may be consistent with a slow conformational change in the enzyme or an unusual change of the ionization state of the catalytic residues. Some evidence has been shown that the protonation state of an iminosugar which is derived from **2** has been observed in the high-resolution structure of Cel5A β-glycosidase in the complex with the inhibitor.[23] Otherwise, based on fluorescence, it was suggested that the slow-onset inhibition of almond β-glycosidase may arise from a conformational change in the enzyme that leads to a high affinity complex.[24] As a result of our simulation, glucosidase occurs little conformational change when the acid/base carboxyl group is protonated in contrast to that when the acid/base carboxyl group is unprotonated, which is based on RMSD analysis. On the other hand, the hydrogen bond interaction between the hydroxyl of the carboxyl group and water plays an important role in the dissociation of the proton in aqueous solution,[36−38] so the hydrogen bond interaction between the acid/base carboxyl group (Glu166) and the solvent was examined and was reported in Figure 9. The number of the hydrogen bond between them is nearly zero. This is perhaps caused by the strong hydrogen bond interaction between the protonated ligand and surrounding water molecules which draws water molecules away from the acid/base carboxyl group. Therefore, the deprotonation of the acid/base carboxyl group would be rather slow. What is more, there are two or three water molecules forming a hydrogen bond with the ligand when the acid/base carboxyl group is protonated, while only a single hydrogen bond is found when the acid/base carboxyl group is unprotonated. This also indicates that the deprotonation may accompany the rearrangement of water. So the deprotonation, accompanying the rearrangement of water should be a slow-onset process and is the cause of slow-onset inhibition of **2** toward the glucosidase. The process of inhibitor binding with glucosidase should probably include two stages. First the conjugate acid of isofagomine binds with the glycosidase when the carboxyl group of acid/base is protonated; afterward the acid/base group occurs deprotonation to achieve more tightly binding, and the later step would be responsible for the observed slow onset of inhibition.

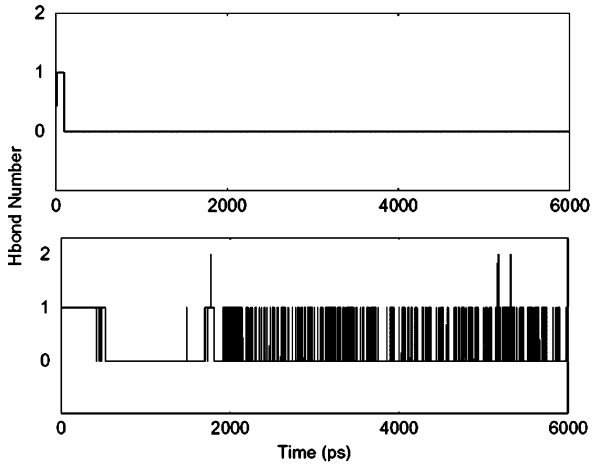


Figure 9. Time dependence of the number of hydrogen bonds between the acid/base carboxyl group (Glu166) and solvent: simulation TmGH-HISO (up) and simulation TmGH-DNJ (down) as comparison.

Thermodynamics of the Binding of Isofagomine and 1-Deoxynojirimycin. It was reported that the binding of both **2** and **1** to β-glucosidase is driven by a large and favorable enthalpy, and the large favorable entropy term makes **2** a better inhibitor than **1**.[18] As shown from the results of our simulation, the large and favorable enthalpy of both binding owes to the strong hydrogen bond and electrostatic interaction between the inhibitor and enzyme, and there is one proton release for both inhibitors when they bind with the enzyme, which is consistent with the result by quantitative analysis of the dependence of ΔH_a on the heat of ionization of the buffer. [18] Shown as hydrogen analysis (Figure 6.A), **2** coordinates only one water molecule which may contribute to the large favorable entropy, while **1** coordinates more than two water molecules (Figure 2.A) together with conformational distortion and may result in an unfavorable entropy. Furthermore, at least one more incorporated water molecule was observed for the binding of **1** with the enzyme relative to that of **2**, which is a reasonable explanation for about nearly −292 J/mol relative difference of heat capacity ($\Delta\Delta C_p$) between the binding of **1** and the binding of **2**.[18,39] The conformation of inhibitor would be largely determined by the interaction between the carboxyl group of the nucleophile and the inhibitor. Shown in Figure 10, as the ligand is 1-deoxynojirimycin, both the imino group and the 2-OH form a strong interaction with the nucleophile carboxyl group, and the ring of the inhibitor is distorted in a skew-boatlike form. While for isofagomine, only the imino group

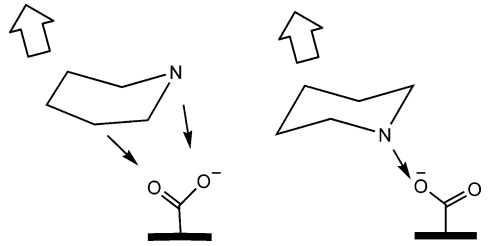


Figure 10. The illumination of the ring conformational of inhibitor induced by the nucleophile carboxyl group: when the inhibitor is 1-deoxynojirimycin (left), when the inhibitor is isofagomine (right), the interaction between the ligand and the nucleophile carboxyl group (black arrow), and the other interaction between ligand and enzyme (white arrow).

forms a strong interaction with the nucleophile carboxyl group, which results in a chairlike form.

Conclusion

1-Deoxynojirimycin and isofagomine are the representations of two sorts of imino-sugar inhibitors which are achieved by the ring oxygen or anomeric carbon of pyranose replaced by the imino group, respectively. These two inhibitors have a distinct thermodynamics property when they bind with β-glucosidase. Nanosecond time scale MD simulations of their complex with β-glucosidase were performed to examine these, and the result is quite reasonable in contrast to the experimental fact. What is more, several interesting conclusions were made and shown as follows:

(1) Just as isofagomine, 1-deoxynojirimycin may bind with the β-glucosidase as a conjugate acid forms according to the comparison of calculated binding free energy and observed binding free energy as well as the comparison of MD snapshot structure and crystal structure.

(2) The slow onset inhibition of isofagomine owns to slow deprotonation of the acid/base carboxyl group (Glu166) and combines with the rearrangement of water in the active site. The nearly zero hydrogen bond interaction between the hydroxyl of the acid/base carboxyl group would be the main cause of slow deprotonation.

(3) The nucleophile carboxyl group (Glu351) plays an important role when the inhibitor binds with glucosidase for it can form a strong hydrogen bond and an electrostatic interaction with both isofagomine and 1-deoxynojirimycin, and such an interaction may determine the ring conformation of the inhibitor.

Acknowledgment. This investigation received financial support from the Innovation Project Foundation of SIOC (Shanghai Institute of Organic Chemistry), and we also thank Dr. Ruo-Wen Wang for his constructive suggest

Supporting Information Available: The atomic charge of the structure HDNJ, DNJ, and HISO as well as the topology and force field parameter file of HDNJ, DNJ, and HISO. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Heightman, T. D.; Vasella, A. T. Recent insights into inhibition, structure, and mechanism of configuration-retaining glycosidase. *Angew. Chem., Int. Ed. Engl.* **1999**, *38*, 750−770.

(2) Zechel, D. L.; Withers, S. G. Glycosidase mechanisms: anatomy of a finely tuned catalyst. *Acc. Chem. Res.* **2000**, *33*, 11−18.

(3) Lillelund, V. H.; Jensen, H. H.; Liang, X.; Bols, M. Recent developments of transition-state analogue glycosidase inhibitors of non-natural product origin. *Chem. Rev.* **2002**, *102*, 515−553.

(4) Kajimoto, T.; Liu, K. K.-C.; Pederson, R. L.; Zhong, Z.; Ichikawa, Y.; Porco, J. A., Jr.; Wong, C.-H. Enzyme-catalyzed aldol condensation for asymmetric synthesis of azasugars: synthesis, evaluation, and modeling of glycosidase inhibitors. *J. Am. Chem. Soc.* **1991**, *113*, 6187−6196.

(5) Dong, W.; Jespersen, T.; Bols, M.; Skrydstrup, T.; Sierks, M. R. Evaluation of isofagomine and its derivatives as potent glycosidase inhibitor. *Biochemistry* **1996**, *35*, 2788−2795.

(6) Ichikawa, Y.; Igarashi, Y.; Ichikawa, M.; Suhara, Y. 1-*N*-Iminosugars: potent and selective inhibitors of β-Glycosidases. *J. Am. Chem. Soc.* **1998**, *120*, 3007−3018.

(7) Kim, Y. J.; Ichikawa, M.; Ichikawa, Y. A rationally designed inhibitor of α-1,3-galactosyltransferase. *J. Am. Chem. Soc.* **1999**, *121*, 5829−5830.

(8) Tanaka, K. S. E.; Winters, G. C.; Batchelor, R. J.; Einstein, F. W. B.; Bennet, A. J. A new structural motif for the design of potent glucosidase inhibitors. *J. Am. Chem. Soc.* **2001**, *123*, 998−999.

(9) Andersen, B.; Rassov, A.; Westergaard, N.; Lundgren, K. Inhibition of glycogenolysis in primary rat hepatocytes by 1,4-dideoxy-1,4-imino-d-arabinitol. *Biochem. J.* **1999**, *342*, 545−550.

(10) Durantel, D.; Branza-Nichita, N.; Carrouee-Durantel, S.; Butters, T. D.; Dwek, R. A.; Zitzmann, N. Study of the mechanism of antiviral action of iminosugar derivatives against bovine viral diarrhea virus. *J. Virol.* **2001**, *75*, 8987−8998.

(11) Goss, P. E.; Baker, M. A.; Carver, J. P.; Dennis, J. W. Inhibitors of carbohydrate processing: A new class of anticancer agents. *Clin. Cancer Res.* **1995**, *1*, 935−944.

(12) Sawkar, A. R.; Cheng, W.-C.; Beutler, E.; Wong, C.-H.; Balch, W. E.; Kelly, J. W. Chemical chaperones increase the cellular activity of N370S beta-glucosidase: A therapeutic strategy for Gaucher disease. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 15428−15433.

(13) Tan, A.; van den Broek, L.; van Boeckel, S.; Ploegh, H.; Bolscher, J. Chemical modification of glucosidase inhibitor 1-Deoxynojirimycin. *J. Biol. Chem.* **1991**, *266*, 14504−14510.

(14) Hempel, A.; Camerman, N.; Mastropaolo, D.; Camerman, A. Glucosidase inhibitor: Structure of deoxynojirimycin and castanospermine. *J. Med. Chem.* **1993**, *36*, 4082−4086.

(15) Bols, M. 1-Aza Sugars, Apparent Transition State Analogues of Equatorial Glycoside Formation/Cleavage. *Acc. Chem. Res.* **1998**, *31*, 1−8.

(16) Bülow, A.; Plesner, I. W.; Bols, M. A Large difference in the thermodynamics of binding of isofagomine and 1-deoxynojirimycin to β-glucosidase. *J. Am. Chem. Soc.* **2000**, *122*, 8567−8568.

(17) Lohse, A.; Hardlei, T.; Jensen, A.; Plesner, I. W.; Bols, M. Investigation of the slow inhibition of almond β-glucosidase and yeast isomaltase by 1-azasugar inhibitors: evidence for the 'direct binding' model. *Biochem. J.* **2000**, *349*, 211−215.

(18) Zechel, D. L.; Boraston, A. B.; Gloster, T.; Boraston, C. M.; Macdonald, J. M.; Tilbrook, D. M. G.; Stick, R. V.; Davies, G. J. Iminosugar Glycosidase Inhibitors: Structural and Thermodynamic Dissection of the Binding of Isofagomine and 1-Deoxynojirimycin to β-Glucosidases. *J. Am. Chem. Soc.* **2003**, *125*, 14313−14323.

(19) Wolfenden, R.; Snider, M. J. The depth of chemical time and the power of enzymes as catalysts. *Acc. Chem. Res.* **2001**, *34*, 938−945.

(20) Karplus, M.; McCammon, J. A. Molecular dynamics simulations of biomolecules. *Nature Struct. Biol*. **2002**, *9* (9), 646−652.

(21) Cardona, F.; Goti, A.; Brandi, A.; Scarselli, M.; Niccolai, N.; Mangani, S. Molecular dynamics simulations on the complexes of glucoamylase II (471) from *Aspergillus awamori* var. X100 with 1-deoxynojirimycin and lentiginosine. *J. Mol. Model.* **1997**, *3*, 249−260.

(22) Dale, M. P.; Ensley, H. E.; Kern, K.; Sastry, K. A.; Byers, L. D. Reversible inhibitors if β-glucosidase. *Biochemistry* **1985**, *24*, 3530−3539.

(23) Legler, G. Glycoside hydrolases: mechanistic information from studies with reversible and irreversible inhibitors. *Adv. Carbohydr. Chem. Biochem.* **1990**, *48*, 319−385.

(24) Varrot, A.; Tarling, C. A.; Macdonald, J. M.; Stick, R. V.; Zechel, D. L.; Withers, S. G.; Davies, G. J. Direct observation of the protonation state of an imino sugar glycosidase inhibitor upon binding. *J. Am. Chem. Soc.* **2003**, *125*, 7496−7497.

(25) 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.

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

(27) (a) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236. (b) Jorgensen, W. L.; McDonald, N. A. *Theochem.* **1998**, *424*, 145−155. (c) Jorgensen, W. L.; McDonald, N. A. *J. Phys. Chem. B* **1998**, *102*, 8049−8059. (d) Rizzo, R. C.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1999**, *121*, 4827−4836. (e) Watkins, E. K.; Jorgensen, W. L. *J. Phys. Chem. A.* **2001**, **105**, 4118−4125.

(28) (a) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Comm.* **1995**, *91*, 43−56. (b) Lindahl, E.; Hess, B.; van der Spoel, D. GROMACS 3.0: A package for molecular simulation and trajectory analysis *J. Mol. Model.* **2001**, *7*, 306−317.

(29) Berendsen, H. J. C.; Postma, J. P. M.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684−3690.

(30) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald potential. *J. Chem. Phys.* **1995**, *103*, 8577−8592.

(31) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463−1472.

(32) Åqvist, J.; Luzhkov, V. B.; Brandsdal, B. O. Ligand binding affinities from MD simulations. *Acc. Chem. Res.* **2002**, *35*, 358−365.

(33) Marelius, J.; Graffner-Nordberg, M.; Hansson, T.; Hallberg, A.; Åqvist, J. Computation of affinity and selectivity: Binding of 2,4-diaminopteridine and 2,4-diaminoquinazoline inhibitors to dihydrofolate reductases. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 119−131.

(34) Hansson, T.; Marelius, J.; Åqvist, J. Ligand Binding Affinity, Prediction by Linear Interaction Energy Methods. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 27−35.

(35) (a) Single-point energies were calculated at the B3LYP/6-311++G(2d, 2p) level. The free energies of solvation in water ΔG_s° were calculated utilizing the AM1-SM5.4 solvation model. (b) Hawkins, G. D.; Giesen, D. J.; Lynch, G. C.; Chambers, C. C.; Rossi, I. J. AMSOL-version 6.8.

(36) Chipot, C.; Gorb, L. C.; Rivail, J.-L. Proton Transfer in the Mono- and the Dihydrated Complexes of HF and HCl: An MP2/6-31+G** ab Initio Study in the Self-Consistent Reaction Field Model of Solvation. *J. Phys. Chem.* **1994**, *98*, 1601−1607.

(37) Smith, A.; Vincent, M. A.; Hillier, I. H. Mechanism of acid dissociation in water clusters: Electronic structure studies of $(H_2O)_nHX$ ($n = 4, 7$; X = OH, F, HS, HSO_3, $OOSO_2H$, $OOH{\cdot}SO_2$). *J. Phys. Chem. A* **1999**, *103*, 1132−1139.

(38) Voegele, A. F.; Klaus, R.; Liedl, K. R. Exploring HBr ionization at the molecular level. *Angew. Chem., Int. Ed.* **2003**, *42*, 2114−2116.

(39) Habermann, S. M.; Murphy, K. P. Energetics of hydrogen bonding in proteins: A model compound study. *Protein Sci.* **1996**, *5*, 1229−1239.

CT050168G

JCTC Journal of Chemical Theory and Computation

All-Atom Calculation of the Normal Modes of Bacteriorhodopsin Using a Sliding Block Iterative Diagonalization Method

Alexey L. Kaledin,* Martina Kaledin,† and Joel M. Bowman

Department of Chemistry and Cherry L. Emerson Center for Scientific Computing, Emory University, Atlanta, Georgia 30322

Received June 22, 2005

Abstract: Conventional normal-mode analysis of molecular vibrations requires computation and storage of the Hessian matrix. For a typical biological system such storage can reach several gigabytes posing difficulties for straightforward implementation. In this work we discuss an iterative block method to carry out full diagonalization of the Hessian while only storing a few vectors in memory. The iterative approach is based on the conjugate gradient formulation of the Davidson algorithm for simultaneous optimization of L roots, where in our case $10 < L < 300$. The procedure is modified further by automatically adding a new vector into the search space for each locked (converged) root and keeping the new vector orthogonal to the eigenvectors previously determined. The higher excited states are then converged with the orthonormality constraint to the locked roots by applying a projector which is carried out using a read-rewind step done once per iteration. This allows for convergence of as many roots as desired without increasing the computer memory. The required Hessian-vector products are calculated *on the fly* as follows, $\mathbf{Kp} = d\mathbf{g}_p/dt$, where $\mathbf{K}$ is the mass weighted Hessian, and $\mathbf{g}_p$ is the gradient along $\mathbf{p}$. The method has been implemented into the TINKER suite of molecular design codes. Preliminary results are presented for the normal modes of bacteriorhodopsin (bR) up to 300 cm^{-1} and for the high frequency range between 2840 and 3680 cm^{-1}. There is evidence of a highly localized, noncollective mode at $\sim$1.4 cm^{-1}, caused by long-range interactions acting between the cytoplasmic and extracellular domains of bR.

1. Introduction

Vibrational modes of proteins are basic motions for protein dynamics and structural transitions. Normal-mode analysis (NMA) is a direct way to analyze vibrational motion.[1] This method has long been used as a tool for interpreting vibrational spectra of small molecules.[2] The frequencies obtained from NMA can be directly related to experimental infrared (IR) and/or Raman measurements. In recent years NMA has been extended to the study of large molecular systems such as proteins.[3−7] Low-frequency modes of proteins are particularly interesting, because they are related to functional properties.[8] It is believed that low-frequency collective modes are responsible for the direct flow of conformational energy in many biological processes.[9−11]

All-atom normal mode calculations of large systems are impeded by the bottleneck associated with computing and storing a full Hessian matrix.[5] An example of this is the 3.6 GB storage required for the Hessian of a system of 10 000 atoms. The use of sparse matrix techniques advocated by some authors can alleviate the storage problem significantly.[12−14] The matrix becomes sparser as the number of atoms increases, and benchmark calculations have been carried out for impressively large nanoparticles.[15] However, it is unclear to what degree the resulting eigenvalues are

* Corresponding author e-mail: akaledi@emory.edu.
† Present address: Department of Chemistry and Biochemistry, Kennesaw State University, Kennesaw, GA 30144.

10.1021/ct050161z CCC: $33.50
Published on Web 11/24/2005

corrupted by removing a large number of small matrix elements arising from the long-range interactions: electrostatic, van der Waals, etc.

Recent advances in Hessian operator techniques[16−19] have opened new possibilities for NMA of macromolecules. Pioneering work of Filippone and Parrinello[16,17] on linear response theory of Hessian demonstrated the use of direct ab initio methods combined with gradients to perform geometry optimization and iterative diagonalization of the Hessian of water dimer without explicit calculation of the Hessian matrix. Using this theory, Reiher and Neugebauer[20−23] did calculations of carbon nanotubes to determine their vibrational modes in the middle range of the spectrum. In a similar spirit, one of us later carried out calculations[24] of up to 200 lowest frequency normal modes of helium nanodroplets with the largest being a 27 000 atom system, affirming applicability of the Hessian operator theory to nanomaterials and, potentially, to biological macromolecules, the latter being the focus of the present work.

The Hessian operator method is ideally suited for combination with iterative diagonalization techniques to solve for the eigenvalues and eigenvectors of very large matrices. Basically, the resultant of Hessian multiplying on a vector is proportional to the gradient change along the vector.[16,17,25−29] This relationship follows from the harmonic expansion of the potential around a given geometry and is summarized by the following expression

$$\mathbf{Kp} = \mathrm{d}\mathbf{g}_{\mathrm{p}}/\mathrm{d}t \tag{1}$$

where $\mathbf{K}$ and $\mathbf{g}_{\mathrm{p}}$ are the mass weighted Hessian and gradient (along vector $\mathbf{p}$). The time derivative on the right-hand side is equivalent to a change over the infinitesimally short trajectory defined by $\mathbf{q} = \pm\delta t\mathbf{p}$, where $\mathbf{q}$ and $\mathbf{p}$ are the mass weighted coordinates and momenta. Thus, given a set of trial vectors spanning a small subspace, a single optimization step can be carried out by computing the residuals and evaluating the appropriate matrix elements using eq 1. The improved vectors are then used to perform another iteration, and so on, until convergence. Clearly, the storage requirement is drastically reduced: from the usual $O(N^2)$ in the conventional calculation with Hessian to $O(N)$ in the iterative calculation using eq 1, where N is the number of atoms.

In the present work we report an extension of a block Davidson iterative method,[30,31] whose modified version was tested in the earlier work,[24] by adding the capability to converge all the normal modes up to a given threshold without compromising the scaling properties of the algorithm. Benchmark calculations are presented for the normal modes of a 222 residue (3503 atoms) protein converged up to 300 cm^{-1} (a total of 1954 normal modes) and between 2840 cm^{-1} and 3680 cm^{-1} (1782 normal modes).

2. Computational Methods

To carry out full diagonalization of the matrix we combine two techniques: (1) a flexible iterative procedure and (2) a memory-efficient evaluation of matrix-vector products.

2.1. Iterative Procedure. The Hessian eigenvalue equation for normal mode I is

$$\mathbf{Ky}_I = \lambda_I\mathbf{y}_I \tag{2}$$

where $\mathbf{y}_I$ is the eigenvector with the corresponding eigenvalue λ_I. The Davidson procedure[30] for finding the lowest root ($I = 1$) of eq 2 involves optimization of a trial vector in an orthogonal subspace, a vector space that is much smaller than the size of the matrix. The approximate solution at iteration n is a linear combination of the n basis vectors, i.e.,

$$\mathbf{y}_I^{(n)} = \sum_i^n c_{Ii}^{(n)}\mathbf{b}_i \tag{3}$$

with the expansion coefficients satisfying the variational condition for the lowest root

$$\mathbf{B}^\dagger\mathbf{KB}\mathbf{c}_I^{(n)} = \lambda_I\mathbf{c}_I^{(n)} \tag{4}$$

where $\mathbf{B}$ is the column matrix of vectors $\mathbf{b}$. The new expansion vector that is added to the iterative subspace $\mathbf{B}$ is derived from perturbation theory,[30] as follows

$$\mathbf{r}_I' = -(\mathbf{D} - \lambda_I\mathbf{1})^{-1}\mathbf{r}_I'' \tag{5}$$

where $\mathbf{D}$ is the diagonal part of $\mathbf{K}$, and $\mathbf{r}_I''$ is the residual of the current approximation to the lowest root, i.e.,

$$\mathbf{r}_I'' = \mathbf{Ky}_I - \lambda_I\mathbf{y}_I \tag{6}$$

One then proceeds by appending the orthogonal complement $\tilde{\mathbf{r}}_I'$ of $\mathbf{r}_I'$ to the subspace $\mathbf{B}$ and diagonalizing the $(n + 1) \times (n + 1)$ interaction matrix $\mathbf{K_{BB}} \equiv \mathbf{B}^\dagger\mathbf{KB}$. The procedure is repeated until the eigenvalue λ_I is stationary (the eigenvalue criterion), or the norm of the residual $\mathbf{r}_I''$ is small enough (the wave function criterion). In case the number of expansions is too large, the procedure is restarted. There exist a variety of methods to improve the diagonal matrix approximation in eq 5,[32] but this discussion is beyond the scope of the present work.

The extension to excited states is straightforward and can be done by simply searching for the next lowest root subject to orthogonality constraint to all the previously converged roots. However, due to the high density of vibrational levels of macromolecules with many weak interactions, it often becomes necessary to perform a simultaneous optimization of several roots. Given a set of trial vectors $\{\mathbf{y}\}_L$ one proceeds by building up the iterative subspace $\mathbf{B}$. Each new iteration expands the subspace by L vectors, where L is the number of roots that are simultaneously optimized. This method is known as the block-Davidson method.[33] Similarly to the single root procedure, the subspace $\mathbf{B}$ is periodically collapsed to L vectors to save space.[31]

In practice, periodic collapse of the $\mathbf{B}$-space to one (or a few) vector per root hinders convergence of the Davidson procedure. Van Lenthe and Pulay[34] first demonstrated on the single root Davidson method that collapsing the $\mathbf{B}$-space on every iteration while retaining the solution vector from the previous iteration basically preserves the variational flexibility of the original method. In other words, the $\mathbf{B}$-space at iteration n consists of three vectors, namely, $\{\tilde{\mathbf{y}}_I^{(n-1)}, \mathbf{y}_I^{(n)}, \tilde{\mathbf{r}}_I^{(n)}\}$, where $\tilde{\mathbf{y}}_I^{(n-1)}$ is the orthogonal complement to $\mathbf{y}_I^{(n)}$, and

$\tilde{\mathbf{r}}_I^{(n)}$ is the *projected* orthogonal complement to $\tilde{\mathbf{y}}_I^{(n-1)}$ and $\mathbf{y}_I^{(n)}$ (cf. eq 8). This method blends together the theory of conjugate gradients and the original Davidson method for the ground state. It has also been shown that simultaneous optimization of several roots for extraction of excited states is possible in this framework.[31,35−37]

The procedure described here is a slight modification of the method suggested by Murray et al.[31] and is a straightforward adaptation of a block version of the van Lenthe−Pulay method[34] for the ground state. The next approximation to root I at iteration n is expanded in a linear combination of orthonormal **B**-space vectors

$$\mathbf{y}_I^{(n+1)} = \sum_j^{3L} c_{Ij}^{(n)}\, \mathbf{b}_j^{(n)} \tag{7}$$

$$= \sum_{j=1}^{L} c_{Ij}^{(n)}\, \mathbf{y}_j^{(n)} + \sum_{j=L+1}^{2L} c_{Ij}^{(n)}\, \tilde{\mathbf{y}}_{j-L}^{(n-1)} + \sum_{j=2L+1}^{3L} c_{Ij}^{(n)}\, \tilde{\mathbf{r}}_{j-2L}^{(n)}$$

where L is the number of roots optimized. On each iteration we solve for the expansion coefficients $\{\boldsymbol{c}_I^{(n)}\}_{3L}$ in eq 7 by diagonalizing the $3L \times 3L$ $\mathbf{K_{BB}}$ matrix. The approximation to the eigenvalue I is the corresponding eigenvalue of $\mathbf{K_{BB}}$. A similar formulation has previously been tested by Knyazev on a number of model problems in physics.[35−37]

With the constraint that the root with index J must be converged before, or simultaneously with, the root of index $J + 1$, etc., the iterations are repeated until the first l roots in the block ($1 \le l \le L$) satisfy certain convergence criteria. (The l converged vectors are then locked and appended to an existing file.) To continue with the iterative process, we use the virtual states at current and previous iteration as the guess for roots $L + 1, \cdots, L + l + 1$. These virtual states are nonoptimized eigenvectors of the $\mathbf{K_{BB}}$ matrix, i.e., $\{\boldsymbol{c}_I^{(n)}\}$, $I = L + 1, \cdots, 3L$, but, as experience shows, they provide an excellent starting point for the upper roots. The procedure does not lose its effectiveness because the virtual states share the conjugate gradient property with the L optimized vectors. The total number of converged roots, designated by a cumulative index M, is increased by l, and the index I is reset to run over the roots $M + 1, \cdots, M + L$. The following iterations simply require that the residuals $\{\mathbf{r}'_I\}_L$ be orthogonal to all the converged vectors (before adding their orthogonal complement to the iterative subspace) which is done by applying the projector to each residual

$$\mathbf{r}_I = \mathbf{r}'_I - \sum_{j=1}^{M} \mathbf{y}_j \mathbf{y}_j^{\dagger} \mathbf{r}'_I \tag{8}$$

The converged vectors $\{\mathbf{y}_I\}_M$ are read from the storage file one at a time and applied successively onto the set $\{\mathbf{r}'_I\}_L$ using eq 8; the file is then rewound to prepare for the next iteration. This read-rewind step is done once per iteration.

The upper extreme of the spectrum can be converged in the same fashion by replacing the Hessian operator with its negative, i.e., $\mathbf{K} \rightarrow -\mathbf{K}$. The eigenvalues change the sign on this transformation, while the eigenvectors are unchanged. Unlike the usual approach of designing the inverse or the shift operator,[5,12] the negative Hessian approach does not require matrix transformations or any additional matrix-vector operations. The high frequency modes are localized and converge significantly faster than the low-frequency ones. This property suggests a useful technique to first converge a bulk of the upper states and then converge the lower ones subject to orthogonality constraint.

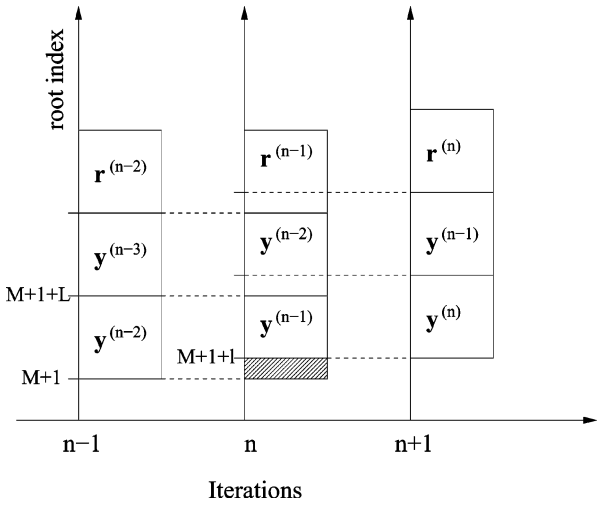


Figure 1. A schematic illustration of the iterative subspace through iterations $n - 1 \rightarrow n \rightarrow n + 1$. It is implied that the first M roots have been converged in $n - 2$ iterations. $\mathbf{y}^{(n)}$ are defined as the vectors obtained by diagonalizing the $3L \times 3L$ matrix at iteration n. In the depicted procedure, at iteration $n - 1$ no new roots have been converged. Thus, the roots $\mathbf{y}_{1,L}^{(n-1)}$, their residuals $\mathbf{r}_{1,L}^{(n-1)}$, and previous roots $\mathbf{y}_{1,L}^{(n-2)}$ are used as the updated basis for iteration n (see the dashed lines). After diagonalization, l lowest roots have been converged at iteration n (shown by the hashed space), and the corresponding vectors are passed on to the next iterations. The procedure can continue until either a frequency threshold is reached or the number of converged roots has reached the desired limit.

The method of projection by eq 8 is similar to the standard deflation techniques, and it achieves the same goal by reducing the search space.[38] However, if many eigenvectors are needed, the convergence criteria must be very strict to ensure that the cumulative error remains small. The overall procedure is stopped after M has exceeded a desired limit, or λ_M has reached a preset threshold. Figure 1 illustrates this process. The above-described procedure can be referred to as a sliding block Davidson−VanLenthe−Pulay method or simply a "sliding block" method. We note that similar iterative techniques based on the Lanczos method[39] have long existed in the literature.[35−37,40−44]

2.2. Hessian-Vector Product. Construction of the $\mathbf{K_{BB}}$ matrix requires evaluation of Hessian-vector products. Given the set of preconditioned and normalized residual vectors we must compute and store their products with the Hessian, $\{\mathbf{K}\tilde{\mathbf{r}}_I\}_L$. It can be shown analytically[16,17] that the algebraic multiplication of the $3N \times 3N$ Hessian matrix on an arbitrary vector is equivalent to differentiation of the gradient along the vector (cf. eq 1). A concise proof of this can be obtained by evaluating the time derivative of the gradient along a classical trajectory, i.e.,

$$\frac{d\mathbf{g}}{dt} = \sum_j \frac{\partial \mathbf{g}}{\partial q_j}\frac{dq_j}{dt} = \mathbf{Kp} \tag{9}$$

where **q** and **p** are the mass weighted coordinates and momenta. Since the trajectory is arbitrary, the momentum **p** can be thought of as an input (trial) vector. The time derivative is evaluated numerically by central differences. Given a unit vector, its product on the Hessian is computed as follows

$$\mathbf{K\hat{u}} = \frac{1}{2\alpha \mathbf{m}^{1/2}}[\nabla V(\mathbf{x} + \delta\mathbf{x}) - \nabla V(\mathbf{x} - \delta\mathbf{x})] \tag{10}$$

where **x** are $3N$ Cartesian coordinates, $\delta\mathbf{x} = \alpha\mathbf{m}^{-1/2}\hat{\mathbf{u}}$, $\alpha = s/(\hat{\mathbf{u}}^\dagger\mathbf{m}^{-1}\hat{\mathbf{u}})^{1/2}$, and **m** is the $3N \times 3N$ diagonal matrix of atomic masses. The displacement parameter s can be chosen in the range $10^{-5}-10^{-3}$ a_0.

2.3. Anharmonicity and Mode Lifetime. Anharmonic effects, such as mode coupling and lifetime, can be estimated directly by computing variations of the eigenvalues of the Hessian. If the normal mode vector $\mathbf{y}_I$ is sufficiently converged, the first derivative of the Hessian expectation value in mode I with respect to the normal coordinates is

$$\nabla \mathbf{y}_I^\dagger \mathbf{K} \mathbf{y}_I = \mathbf{y}_I^\dagger (\nabla \mathbf{K}) \mathbf{y}_I \tag{11}$$

The right-hand side contains the first derivative of the Hessian which carries the information of third derivatives of the potential. Similar to eq 10 the differentiation is done numerically. For the normal coordinate J

$$\left(\frac{\partial \mathbf{K}}{\partial Q_J}\right)\mathbf{y}_I = \frac{1}{2\delta Q_J}[\mathbf{K}(\mathbf{x} + \delta\mathbf{x}_J)\mathbf{y}_I - \mathbf{K}(\mathbf{x} - \delta\mathbf{x}_J)\mathbf{y}_I] \tag{12}$$

where $\delta\mathbf{x}_J$ is the Cartesian displacement vector along the normal direction J. The two Hessian vector products are then evaluated using eq 10 resulting in a total of *four* gradient computations. Using $\partial\omega_I/\partial Q_J = (\partial\lambda_I/\partial Q_J)/(2\omega_I)$ and $\delta Q_J \equiv \alpha_J$, we obtain the following expression for the derivative of the frequency

$$\frac{\partial \omega_I}{\partial Q_J} = \frac{1}{8\alpha_I\alpha_J\omega_I}\mathbf{y}_I^\dagger \mathbf{m}^{-1/2}(\nabla \mathrm{V}(\mathbf{x} + \delta\mathbf{x}_{IJ}^s) + \nabla \mathrm{V}(\mathbf{x} - \delta\mathbf{x}_{IJ}^s) - \nabla \mathrm{V}(\mathbf{x} + \delta\mathbf{x}_{IJ}^d) - \nabla \mathrm{V}(\mathbf{x} - \delta\mathbf{x}_{IJ}^d)) \tag{13}$$

where $\delta\mathbf{x}_{IJ}^{s/d} \equiv \delta\mathbf{x}_J \pm \delta\mathbf{x}_I$. For $I = J$, eq 13 provides a measure of anharmonicity of mode I, while for $I \neq J$ it yields two-mode coupling strength.

The first derivatives can be used to calculate fluctuation of frequencies and consequently the lifetime of a particular mode. The quantum mechanical expression for the variance of the frequency of normal mode I is

$$\langle \Delta\omega_I^2 \rangle = \langle \omega_I^2(\mathbf{Q}) \rangle - \langle \omega_I(\mathbf{Q}) \rangle^2 \tag{14}$$

where brackets imply a thermal average over **Q**. Using eq 13 to expand the frequency to the first order in $\delta\mathbf{Q} \equiv \mathbf{Q} - \mathbf{Q}_{eq}$ and after the cancellation of the linear and the constant terms we obtain a simplified result

$$\langle \Delta\omega_I^2 \rangle = \langle \delta\mathbf{Q}^\dagger \Omega_I \delta\mathbf{Q} \rangle \tag{15}$$

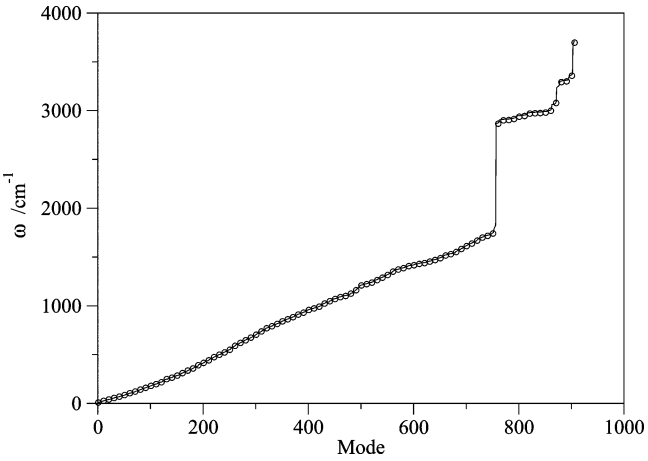


Figure 2. Frequencies of Trp-cage obtained with the block Davidson method (open circle, only every 10th frequency is show for clarity). The exact frequencies from full diagonalization (solid line) are shown for comparison.

where Ω_I is the tensor of first derivative moments, and the notation ω_I^J is short form for derivative with respect to J. The integration is completed analytically in the normal mode basis

$$\langle \Delta\omega_I^2 \rangle_{QM} = \frac{\hbar}{2}\sum_{J=1}^{M}\frac{(\omega_I^J)^2}{\omega_J}\coth\left(\frac{\hbar\omega_J}{2k_BT}\right) \tag{16}$$

The summation runs over the available normal modes. From the uncertainty principle, the lifetime of a mode can be estimated as $\tau_I \sim 1/\langle \Delta\omega_I^2 \rangle^{1/2}$. Equation 16 should provide a reliable estimate for lifetimes at low temperatures where the cubic terms in the potential dominate nonharmonic dynamics. A more rigorous theory for calculation of the lifetime involves a quantum mechanical treatment using perturbation theory,[45] as has been applied in similar calculations.[46,47]

The present method to estimate lifetime is closely related to the classical molecular dynamics and Monte Carlo simulation of the $\langle \Delta\omega^2 \rangle$ quantity, where the averaging is done over phase space classically. The corresponding classical counterpart of expression 16 in the $\hbar \rightarrow 0$ limit yields

$$\langle \Delta\omega_I^2 \rangle_{CM} = k_BT\sum_{J=1}^{M}\left(\frac{\omega_I^J}{\omega_J}\right)^2 \tag{17}$$

The results for $\langle \Delta\omega^2 \rangle$ bear close similarity to the well-known NMA expressions for atomic square fluctuations.[5]

3. Vibrational Modes of Bacteriorhodopsin

We implemented the sliding block iterative diagonalization method into the TINKER[48] suite of molecular modeling codes. The method was tested first on a small protein, Trp-cage (PDB code: 1L2Y,[49] 20 residues) for which exact normal-mode frequencies can be calculated using standard matrix diagonalization with the explicit Hessian matrix. The potential function of Trp-cage was described with the AMBER force field *ff98* for nucleic acids.[50,51] The protein was first energy minimized until the RMS gradient was less than 10^{-6} kcal/(mol Å). Figure 2 shows Trp-cage frequencies

obtained by the two methods. The two sets of eigenvectors were compared by calculating their overlap, which on average was 99.999%.

We now turn to the much larger protein, for which the calculation and storage of the full Hessian is prohibitive. Bacteriorhodopsin is a transmembrane protein found in the purple membrane of Halobacterium salinarum.[52] The study of bR has become an area of considerable interest in biochemistry seeking information about the protein's dynamics and function, for three main reasons.[53] The protein is unusually stable. It exhibits strong spectral shifts in the 400−600 nm range which are connected to reaction intermediates, and it is possible to measure vibrational spectra, characterizing geometries as well as protonated states.

The structure, dynamics, and energetics of bR have been studied extensively by molecular dynamics simulations.[54−58] Conformational modes of bR have also been studied using inelastic neutron scattering.[59,60] Recently, far-infrared (FIR) spectral measurements of wild-type (WT) and D96N mutant bR have been carried out using terahertz time domain spectroscopy.[61] In the same work,[61] the lowest few normal modes of bR were calculated using the iterative diagonalization method of Mouawad and Perahia[62] and compared to the experimental measurements. Those calculations revealed the lowest frequency mode at ∼10 cm^{-1}. Some very low-frequency modes (below 10 cm^{-1}) observed experimentally were missing in this theoretical spectrum. We noted that in this normal mode calculation[61] strict cutoffs were imposed for the nonbonded interactions.

In the present calculations, geometry optimization and the NMA were carried out in the gas phase without any cutoffs imposed on the long-range interactions. Previous studies,[63] for example, pointed out the existence of long-range interactions between the cytoplasmic and extracellular surface domains of bR that are mediated by salt bridges and hydrogen-bonded networks. Such long-range interactions are therefore expected to be of functional significance. The X-ray diffraction structure of WT-bR (PDB code: 1C3W[64]) served as the starting point for geometry optimization. The potential function was described with the Charmm27 parameter set.[65,66] The structure was energy minimized until the RMS gradient was less than 10^{-5} kcal/(mol Å). Full normal mode calculation of WT-bR would require ∼0.4 GB of memory, while the present method required a maximum of 6.3 MB. A convergence criterion of 0.001 cm^{-1} for the frequency was used for all calculations.

The first 1954 normal modes up to 300 cm^{-1} were calculated in four stages, 0−100, 100−200, 200−250, 250−300 cm^{-1}. To converge the first 696 normal modes up to 100 cm^{-1} we used a 200-vector block starting with a random set of vectors. The procedure required 807 iterations and took ∼300 h of CPU time on a single 2.4 GHz processor. The other three stages were completed with 100-vector blocks. It was observed that the lowest root in each stage was always higher than the highest root of the previous stage, as is required by the variational principle. We are thus confident that no roots were missed in the procedure. Figure 3 shows the convergence profile of the lowest 10 normal modes in the later steps of diagonalization. The early steps of the algorithm, iterations 1−20, quickly remove the high frequency components from the guess vectors. The following iterations simply work to refine the strongly coupled vectors, and it may take hundreds of iterations to cleanly separate the true eigenstates. Thus, the initial guess is not as important as the size of the block (the bigger the block, the more efficient the convergence) or the preconditioning scheme. Reiher et al. investigated the effects of the approximate inverse of **K** in eq 5 and found encouraging results.[20] Their scheme can also be applied in the present calculations.

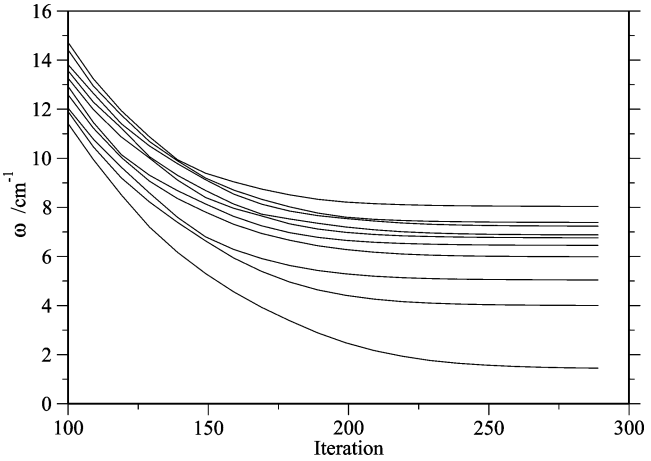


Figure 3. The convergence of the lowest 10 normal-mode frequencies of WT-bR. Note the configuration mixing occurring between 100 and 200 iterations, seen here as "avoided crossing". The nearly degenerate pairs (6,7) and (8,9) change character several times before iteration 200. The lowest root uncouples from the rest at early stages but converges very slowly.

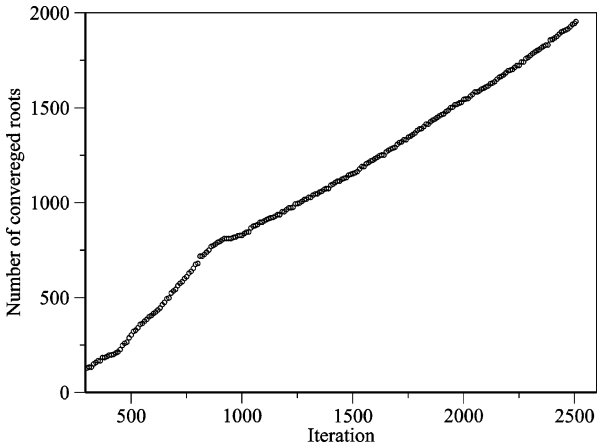


Figure 4. The number of converged normal-mode frequencies of WT-bR as a function of iterations. Note the crossover point at 807 iteration where the block size was reduced from 200 to 100.

Figure 4 demonstrates the dependence of the number of converged roots as a function of iterations. The lowest part of the spectrum that contains many delocalized vibrations is very difficult to converge. Note that it took 294 iterations to converge the first root, $\omega_1 = 1.442$ cm^{-1}. Overall, the first 123 modes converged in 300 iterations. The convergence curve as a function of iteration appears to be a superposition of two lines. The crossover point occurs at 807 iteration

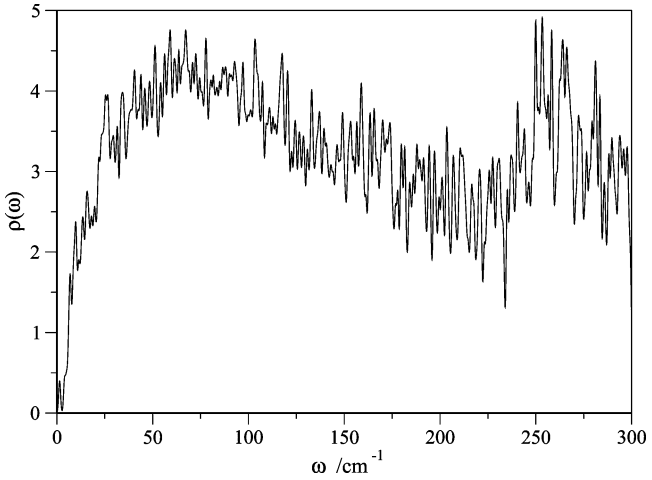


Figure 5. Unaveraged density of normal modes of WT-bR.

where the block size was changed from $L = 200$ to $L = 100$. The smaller block results in more iterations per root. Simple extrapolation can give an estimate of the computational cost required to obtain more roots, provided the density of states does not change rapidly.

A density of the normal modes of WT-bR up to 300 cm^{-1} is plotted in Figure 5. The NM distribution has been represented as a sum of Gaussians with a width 0.5 cm^{-1}. The density of the states is very broad and increasing up to 75 cm^{-1}, similar to experimental measurements.[61] Above 75 cm^{-1} the density of the states slowly decreases and then goes up again at 250 cm^{-1}. Experimental measurements of infrared absorbance of bR in solution show similar behavior.[11]

Among many useful properties, normal modes can be used to determine the role of collective motions in the dynamics of the system. The participation ratios have been used to characterize the degree of delocalization of the normal modes in liquid[67] and protein systems.[69,70] The participation ratios are defined as follows

$$R_I^a = \sum_j^{3N} (y_{Ij})^4 \tag{18}$$

$$R_I^r = \sum_l^{N_r} [\sum_j^{3N_l} (y_{Ij})^2]^2$$

where N_r is the number of residues, and N_l is the number of atoms in the lth residue. One can interpret $1/R_I^a$ as the number of degrees of freedom involved in the Ith mode and $1/R_I^r$ as the number of protein residues participating in that mode. If a mode is completely localized, only one of the eigenvector coefficients will be nonzero and $1/R_I^a$ will be equal to unity. On the other hand, if a mode is completely delocalized, each degree of freedom will be equally involved in that mode and $1/R_I^a$ will be equal to $3N$.

Table 1 shows the lowest 20 normal-mode frequencies of WT-bR up to 10 cm^{-1} with the corresponding residual norms (eq 6), quantum lifetimes (eq 16), and participation ratios (eq 18). Low-frequency modes are typically delocalized throughout the protein and involve mainly collective movements of residues. Most of the normal modes of WT-bR up to 10 cm^{-1} are delocalized, with $1/R_I^a > 200$. However, the lowest normal mode $\omega_1 = 1.442$ cm^{-1} is almost completely localized on the loop that connects helices B and C (Figure 6), and the participation ratio suggests involvement of only 5 residues (residues 68−72) GLY-GLY-GLU-GLN-ASN.

Table 1. Lowest Frequencies in cm^{-1} of WT-bR with the Associated Residual Errors, Quantum Lifetime in ps, and the Inverse Participation Ratios Defined in Eq 18

I	ω_I	$\|r_I\|$	τ_I(0 K)	$1/R_I^a$	$1/R_I^r$
1	1.442	0.18156E-8	0.16	39	5
2	4.003	0.13395E-8	11.0	1211	99
3	5.042	0.12619E-8	5.0	894	89
4	5.988	0.15719E-8	8.6	993	86
5	6.452	0.12818E-8	15.1	306	24
6	6.760	0.14046E-8	11.5	914	75
7	6.877	0.18801E-8	6.8	382	37
8	7.234	0.19270E-8	6.8	284	29
9	7.388	0.13283E-8	9.5	79	8
10	8.040	0.16267E-8	21.4	857	64
11	8.434	0.11833E-8	6.8	600	56
12	8.530	0.16593E-8	20.0	807	50
13	8.883	0.16471E-8	4.1	388	42
14	9.088	0.21947E-8	22.2	238	21
15	9.281	0.15329E-8	18.5	325	31
16	9.737	0.18148E-8	24.5	236	26
17	9.778	0.13752E-8	12.6	353	30
18	9.959	0.14946E-8	24.4	817	68
19	10.111	0.16286E-8	15.3	157	14
20	10.190	0.16805E-8	17.4	900	76

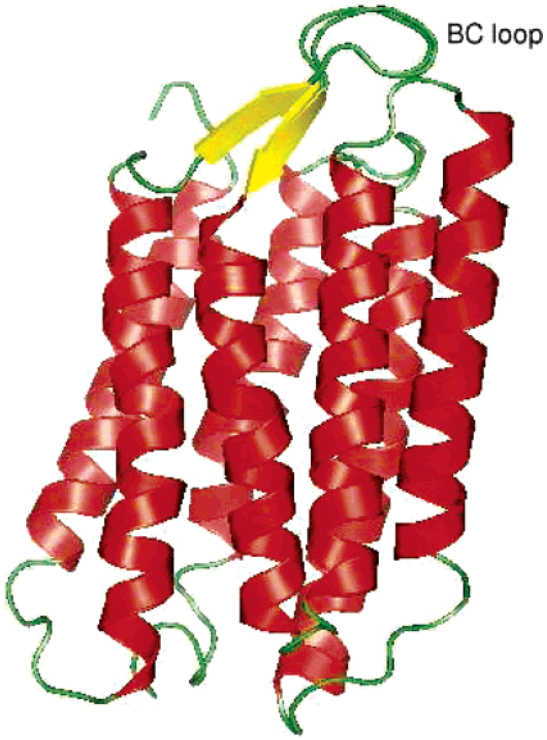


Figure 6. The 1.442 cm^{-1} mode of WT-bR represented as two superimposed structures: the equilibrium and the slightly displaced structure with $\delta E = 3.6$ cm^{-1} along the normal mode vector. The largest displacements occur in the BC loop in the extracellular part of the protein. The figure was created with PyMOL.[68]

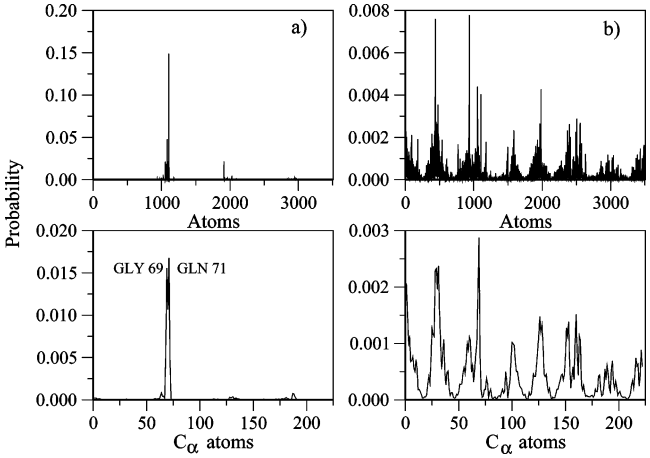


Figure 7. Squared amplitudes summed over the three Cartesian directions for all atoms (upper panels) and C_α atoms (bottom panels) for the lowest two normal modes (a) $\omega_1 = 1.442$ cm^{-1} and (b) $\omega_2 = 4.003$ cm^{-1}. The lowest normal mode is completely localized on the BC loop, while the second mode represents typical collective motion.

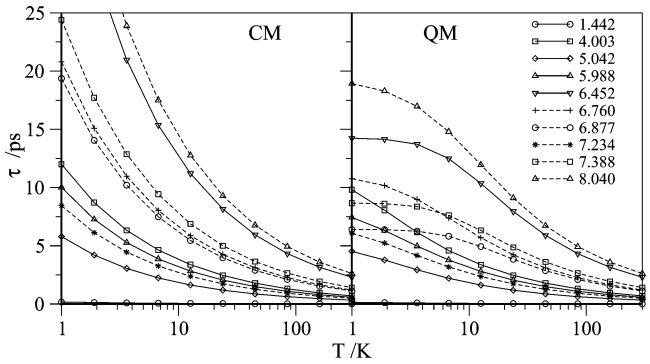


Figure 8. Temperature dependence of lifetimes up to 300 K of several low-frequency modes of WT-bR. The lifetime calculations were done using formulas 16 (for QM results) and 17 (for CM results). The summations included the states up to 40 cm^{-1}, a total of 200 states. The horizontal scale is logarithm base of 10.

For comparison, we plot displacements of all atoms and C_α's of the two lowest normal modes in Figure 7. The largest displacements correspond to C_α's of GLY-69 and GLN-71 residues. We noted that the C_α's displacements are much smaller than displacements of the protein side chains. The greater flexibility of the interhelical loops was predicted in molecular dynamics studies.[71,72]

The lifetime of the highly localized lowest frequency mode, $\tau_1 \approx 161$ fs, is an order of magnitude shorter than the lifetime of the other modes, pointing to its spatial instability, i.e., propensity to jump to another minimum. Thermal stability of the normal modes is depicted in Figure 8 where we compare quantum and classical calculations. As expected, the lower frequency modes reach the classical limit ($1/k_BT \rightarrow 0$) faster than the higher frequency ones, and, already at 10 K, the classical results for all the frequencies up to 10 cm^{-1} are similar to the quantum ones. It is interesting to note that some modes destabilize much quicker than others with increasing temperature. For example, $\omega_2 = 4.003$ cm^{-1} which corresponds to the collective motion (see Figure 7b)

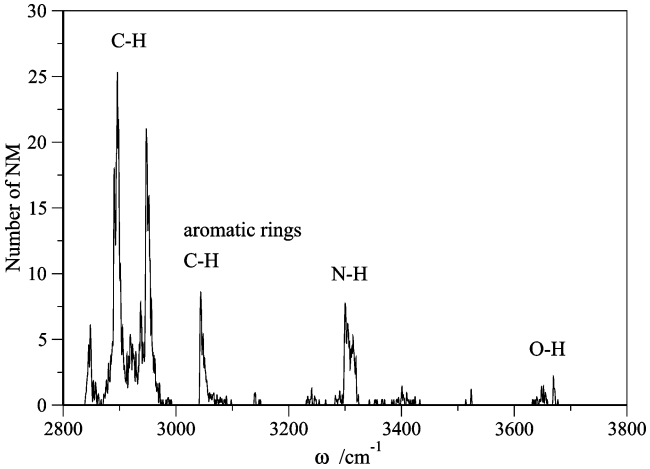


Figure 9. High-frequency vibrations of WT-bR.

has a sharply decreasing lifetime curve, crossing two states. On the other hand, $\omega_7 = 6.877$ cm^{-1} has a flat lifetime curve up to 10 K.

The upper extreme of the spectrum between 2840 and 3680 cm^{-1} was also obtained using the technique mentioned in section 2. For the dominant spectral features the atomic motion was analyzed, and the peaks were assigned (Figure 9). The spectrum contains localized A−H vibrations (where A = C, N, O) that converged more rapidly than the collective modes. Using a 100 vector block, it took 700 iterations to converge all the 1782 high frequency modes. The sliding block procedure clearly identified the gap in the density of states of WT-bR. The modes $\omega_{8721} = 1751.8$ cm^{-1} localized on the E helix (C−C stretch of the TRP-134 side chain) and $\omega_{8722} = 2839.3$ cm^{-1} localized on the BC loop (C−H stretch of the MET-64 side chain) represent the left and right sides of the gap in the spectrum.

To briefly address the usefulness of sparse matrix diagonalization approaches and possibly shed light on the origin of the localized lowest frequency mode, we performed diagonalization using a cutoff scheme for the long-range interactions. All the nonbonded interactions were truncated at several values, and the lowest few roots were converged for each value of the cutoff radius, r_c. Truncation of the nonbonded interactions is equivalent to the removal of small off-diagonal Hessian elements based on a threshold. The resulting eigenvalues are approximate, although the errors of such calculations are rarely reported. Figure 10 shows the dependence of the first three frequencies on the cutoff radius. The frequencies are barely perturbed for $r_c > 25$ Å indicating that there are no dynamically significant interactions beyond 25 Å. However, as the cutoff radius is made smaller, the lowest mode undergoes substantial variations in frequency, and in the range $12 < r_c < 22$ Å it becomes unstable, while the excited states remain roughly the same. If the nonbonded forces are removed for all distances less than 10−11 Å, the modes lose their identity, as seen by the plunging curves in the figure, and the protein is possibly distorted to a nearby structure (a local minimum). The participation ratios of the lowest mode are also quite sensitive to the interaction radius. At 15 Å, for example, $1/R_I$ is about 84 and still bears characteristics of a localized mode, but at 11 Å, R_I is 1100, the signature of a typical collective mode.

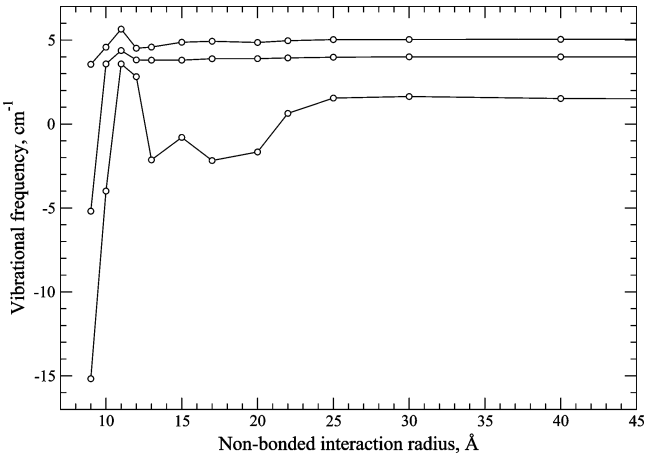


Figure 10. Dependence of the lowest three modes on the nonbonded interaction radius. The negative values on the vertical axis designate imaginary frequencies.

4. Concluding Remarks

The gradient based iterative diagonalization method presented here is designed specifically to treat large biomolecules on the all-atom basis. The method is CPU-bound, rather than memory-bound, and its scaling properties are akin to molecular dynamics simulations. Each iteration L Hessian-vector products ($2L$ gradients) are computed. In sequential calculations the gradient evaluations comprise the main bottleneck, although it is clear that massive parallelization (for large L) is almost trivial since the L Hessian-vector products are unrelated to each other.

Because the iterative subspace is nonexpanding, only the $3L$ vectors and their $3L$ Hessian products need to be stored in memory, resulting in a total storage of $18LN$ elements for a molecule with N atoms. In double precision, the rule of thumb for the memory is 144 MB for a 100-vector block per 10^4 atoms. Thus, all-atom calculations for systems with up to 10^6 atoms are not unrealistic, and it is possible to converge the entire spectrum while storing only the $3L$ vectors and their $3L$ Hessian products in memory.

Calculations are presented for a moderate size protein, WT-bR (3503 protein atoms, 222 residues), in gas phase. Analysis of the results points to two observations: (i) the existence of noncollective, localized low-frequency modes in proteins, also found in the recent calculations of Cui et al.[73] and (ii) the importance of the long-range electrostatic and van der Waals interactions. In conventional macromolecular simulations, to speed up the calculation, cutoff schemes[74] are applied to all nonbonded interactions. (For example, in WT-bR the nonbonded interactions cutoff at 20 and 13 Å reduced the effort to compute the gradient by three and four times, respectively.) Although the cutoff approximation may not always cause severe problems, it should be used with caution, especially if the simulation involves charge transfer. The calculations presented here have revealed that long-range forces acting up to 25 Å can destabilize some of the low-frequency modes. It is thus conceivable that the NMA methods that use sparse matrix techniques may not properly capture the vibrational dynamics by simply discarding the many small matrix elements. More numerical tests are necessary to better understand the potential dangers of using the sparse Hessian approaches for macromolecules.

Acknowledgment. We thank the National Science Foundation for the support, grant No. CHE-0446527.

References

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

(2) Herzberg, G. *Molecular Spectra and Molecular Structure. II. Infrared and Raman Spectra of Polyatomic Molecules*; Van Nostrand: New York, 1945.

(3) *Normal mode analysis: Theory and applications to biological and chemical systems*; Cui, Q., Bahar, I., Eds.; CRC Press: 2005.

(4) Brooks, B. R.; Janežič, D.; Karplus, M. *J. Comput. Chem.* **1995**, *16*, 1522−1542.

(5) Janežič, D.; Brooks, B. R. *J. Comput. Chem.* **1995**, *12*, 1543−1553.

(6) Marques, O. A.; Sanejouand, Y.-H. *Proteins* **1995**, *23*, 557−560.

(7) Tama, F.; Sanejouand, Y.-H. *Protein Eng.* **2001**, *14*, 1−6.

(8) Durand, P.; Trinquier, G.; Sanejouand, Y.-H. *Biopolymers* **1994**, *34*, 759−771.

(9) Austin, R. H.; Hong, M. K.; Moser, C.; Plombon, J. *Chem. Phys.* **1991**, *158*, 473−486.

(10) Beratan, D. N.; Betts, J. N.; Onuchic, J. N. *J. Phys. Chem.* **1992**, *96*, 2852−2855.

(11) Xie, A.; Van der Meer, A. F. G.; Austin, R. H. *Phys. Rev. Lett.* **2002**, *88*, 018102-1−018102-4.

(12) Noid, D. W.; Fukui, K.; Sumpter, B. G.; Yang, C.; Tuzun, R. E. *Chem. Phys. Lett.* **2000**, *316*, 285−296.

(13) Fukui, K.; Sumpter, B. G.; Noid, D. W.; Yang, C.; Tuzun, R. E. *Comput. Theor. Polym. Sci.* **2001**, *11*, 191−196.

(14) Fukui, K.; Sumpter, B. G.; Noid, D. W.; Yang, C.; Tuzun, R. E. *J. Phys. Chem. B* **2000**, *104*, 526−531.

(15) Noid, D. W. private communication.

(16) Filippone, F.; Parrinello, M. *Chem. Phys. Lett.* **2001**, *345*, 179−182.

(17) Filippone, F.; Meloni, S.; Parrinello, M. *J. Chem. Phys.* **2001**, *115*, 636−642.

(18) Anglada, J. P.; Besalù E.; Bofill, J. M.; Rubio, J. *J. Math. Chem.* **1999**, *25*, 85−92.

(19) Prat-Resina, X.; Garcia-Viloca, M.; Monard, G.; González A.; Lluch, J. M.; Bofill, J. M.; Anglada, J. M. *Theor. Chem. Acc.* **2002**, *107*, 147−153.

(20) Reiher, M.; Neugebauer, J. *J. Chem. Phys.* **2003**, *118*, 1634−1641.

(21) Reiher, M.; Neugebauer, J. *Phys. Chem. Chem. Phys.* **2004**, *6*, 4621−4629.

(22) Neugebauer, J.; Reiher, M. *J. Comput. Chem.* **2004**, *25*, 587−597.

(23) Neugebauer, J.; Reiher, M. *J. Phys. Chem. A* **2004**, *108*, 2053−2061.

(24) Kaledin, A. L. *J. Chem. Phys.* **2005**, *122*, 184106-1−184106-7.

(25) Gerratt, J.; Mills, I. M. *J. Chem. Phys.* **1968**, *49*, 1719−1729.

(26) Pick, R. M.; Cohen, M. H.; Martin, R. M. *Phys. Rev. B* **1970**, *1*, 910.

(27) DeCicco, P. D.; Johnson, F. A. *Proc. R. Soc. A* **1969**, *310*, 111−119.

(28) Baroni, S.; Giannozzi, P.; Testa, A. *Phys. Rev. Lett.* **1987**, *59*, 2662.

(29) Gonze, X.; Vigneron, J.-P. *Phys. Rev. B* **1989**, *39*, 13120−13128.

(30) Davidson, E. R. *J. Comput. Phys.* **1975**, *17*, 87.

(31) Murray, C. W.; Racine, S. C.; Davidson, E. R. *J. Comput. Phys.* **1992**, *103*, 382−389.

(32) Leininger, M. L.; Sherrill, C. D.; Allen, W. D.; Schaefer, H. F., III *J. Comput. Chem.* **2001**, *22*, 1574−1589.

(33) Liu, B. *Numerical Algorithms in Chemistry: Algebraic Methods*; Moler, C., Shavitt, I., Eds.; LBL-8158; Lawrence Berkeley Laboratory: Berkeley, CA, 1978.

(34) Van Lenthe, J. H.; Pulay, P. *J. Comput. Chem.* **1990**, *11*, 1164−1168.

(35) Knyazev, A. V. *Electron. Trans. Num. Anal.* **1998**, *7*, 104−123.

(36) Knyazev, A. V. *Int. Ser. Numer. Math.* **1991**, *96*, 143−154.

(37) Knyazev, A. V. *SIAM J. Sci. Comput.* **2001**, *23*, 517−541.

(38) Saad, Y. *Numerical methods for large eigenvalue problems*; Manchester University Press: Halsted Press: Wiley: 1992.

(39) Lanczos, C. *J. Res. Nat. Bur. Standards* **1950**, *45*, 255−282.

(40) Sleijpen, G. L. G.; Van der Vorst, H. A. *SIAM Rev.* **2000**, *42*, 267−29.

(41) Lehoucq, R.; Sorensen, D. *SIAM J. Mater. Anal. Appl.* **1996**, *8*, 789−821.

(42) Lehoucq, R. B.; Sorensen, D. C.; Yang, C. *ARPACK Users' Guide: Solution of Large-Scale Eigenvalue Problems with Implicitly Restarted Arnoldi Methods*; SIAM: Philadelphia, PA, 1998.

(43) Anglada, J. M.; Besalù E.; Bofill, J. M. *Theor. Chem. Acc.* **1999**, *103*, 163−166.

(44) Bofill, J. M.; Moreira, I. D. P. R.; Anglada, J. M.; Illas, F. *J. Comput. Chem.* **2000**, *21*, 1375−1386.

(45) Maradudin, A. A.; Fein, A. E. *Phys. Rev.* **1962**, *128*, 2589−2608.

(46) Yu, X.; Leitner, D. M. *J. Phys. Chem. B* **2003**, *107*, 1698−1707.

(47) Fujisaki, H.; Bu, L.; Straub, J. E. *Adv. Chem. Phys.* **2005**, *130*, 179−203.

(48) Ponder, J. TINKER Software Tools for Molecular Design, Version 4.2, Washington University School of Medicine, June 2004, available from http:/dasher.wustl.edu/tinker.

(49) Neidigh, J. W.; Fesinmeyer, R. M.; Andersen, N. H. *Nature Struct. Biol.* **2002**, *9*, 425−430.

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

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

(52) Lanyi, J. K. *Int. Rev. Cytol.* **1999**, *187*, 161−202.

(53) Ben-Nun, M.; Molnar, F.; Lu, H.; Phillips, J. C.; Martinez, T. J.; Schulten, K. *Faraday Discuss.* **1998**, *110*, 447−462.

(54) Ferrand, M.; Dianoux, A. J.; Petry, W.; Zaccai, G. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 9668−9672.

(55) Humphrey, W.; Xu, D.; Sheves, M.; Schulten, K. *J. Phys. Chem.* **1995**, *99*, 14549−14560.

(56) Xu, D.; Martin, C.; Schulten. K. *Biophys. J.* **1996**, *70*, 453−460.

(57) Xu, D.; Sheves, M.; Schulten, K. *Biophys. J.* **1995**, *69*, 2745−2760.

(58) Roux, B.; Nina, M.; Pomes, R.; Smith, J. C. *Biophys. J.* **1996**, *71*, 670−681.

(59) Ferrand, M.; Zaccai, G.; Nina, M.; Smith, J. C.; Etchebest, C.; Roux, B. *FEBS. Lett.* **1993**, *327*, 256−260.

(60) Zaccai, G. *Science* **2000**, *288*, 1604−1607.

(61) Whitmire, S. E.; Wolpert, D.; Markelz, A. G.; Hillbrecht, J. R.; Galan, J.; Birge, R. R. *Biophys. J.* **2003**, *85*, 1269−1277.

(62) Mouawad, L.; Perahia, D. *Biopolymers* **1996**, *33*, 599−611.

(63) Alexiev, U.; Mollaaghababa, R.; Khorana, H. G.; Heyn, M. P. *J. Biol. Chem.* **2000**, *275*, 13431−13440.

(64) Luecke, H.; Schobert, B.; Richter, H. T.; Cartailler, J. P.; Lanyi, J. K. *J. Mol. Biol.* **1999**, *291*, 899−911.

(65) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., 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. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.

(66) Foloppe, N.; MacKerell, A. D., Jr. *J. Comput. Chem.* **2000**, *21*, 86−104.

(67) Cho, M.; Fleming, G. R.; Saito, S.; Ohmine, I.; Stratt, R. M. *J. Chem. Phys.* **1994**, *100*, 6672−6683.

(68) DeLano, W. L. The PyMOL Molecular Graphics System, DeLano Scientific LLC, San Carlos, CA. http://www.pymol.org.

(69) Sagnella, D. E.; Straub, J. E. *Biophys. J.* **1999**, *77*, 70−84.

(70) Bu, L.; Straub, J. E. *Biophys. J.* **2003**, *85*, 1429−1439.

(71) Rajamani, R.; Gao, J. *J. Comput. Chem.* **2002**, *23*, 96−104.

(72) Jang, H.; Crozier, P. S.; Stevens, M. J.; Woolf, T. B. *Biophys. J.* **2004**, *87*, 129−145.

(73) Cui, Q.; Li, G.; Ma, J.; Karplus. M. *J. Mol. Biol.* **2004**, *340*, 345−372.

(74) Steinbach, P. J.; Brooks, B. R. *J. Comput. Chem.* **1994**, *15*, 667−683.

CT050161Z

JCTC Journal of Chemical Theory and Computation

A Self-Consistent Space-Domain Decomposition Method for QM/MM Computations of Protein Electrostatic Potentials

Jose A. Gascon,[†] Siegfried S. F. Leung,[†] Enrique R. Batista,[‡] and Victor S. Batista*,[†]

Department of Chemistry, Yale University, P.O. Box 208107, New Haven, Connecticut 06520-8107, and Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545

Received September 4, 2005

Abstract: This paper introduces a self-consistent computational protocol for modeling protein electrostatic potentials according to static point-charge model distributions. The protocol involves a simple space-domain decomposition scheme where individual molecular domains are modeled as Quantum-Mechanical (QM) layers embedded in the otherwise classical Molecular-Mechanics (MM) protein environment. ElectroStatic-Potential (ESP) atomic charges of the constituent molecular domains are computed, to account for mutual polarization effects, and iterated until obtaining a self-consistent point-charge model of the protein electrostatic potential. The novel protocol achieves quantitative agreement with full QM calculations in the description of electrostatic potentials of small polypeptides where polarization effects are significant, showing a remarkable improvement relative to the corresponding electrostatic potentials obtained with popular MM force fields. The capabilities of the method are demonstrated in several applications, including calculations of the electrostatic potential in the potassium channel protein and the description of protein−protein electrostatic interactions.

1. Introduction

The development of rigorous and practical methods for the accurate description of molecular electrostatic potentials is a subject of great interest,[1−25] since the energetics of molecular processes is often dominated by electrostatic energy contributions.[26−52] In particular, electrostatic interactions play a central role in a variety of molecular processes in biological molecules,[26−29] including enzyme catalysis,[30,31] electron transfer,[32,33] proton transport,[30,34−37] ion channels,[38,39] docking and ligand binding,[40−45] macromolecular assembly,[46−50] and signal transduction.[51,52] However, a rigorous and practical ab initio method to compute accurate electrostatic potentials of biological molecules has yet to be established.[53−59] This paper introduces one such method, an approach to obtain static point-charge models of protein electrostatic potentials by combining a novel iterative self-consistent space-domain decomposition scheme with conventional Quantum Mechanics/Molecular Mechanics (QM/MM) hybrid methods.

QM/MM hybrid methods partition the system into QM and MM layers,[60] offering an ideally suited approach for describing the polarization of a molecular domain due to the influence of the surrounding (protein) environment. Such a methodology models the electrostatic perturbation of the MM layer, on the QM domain, according to the static point-charge model distributions prescribed by MM force fields.[61−66] However, it is widely recognized that standard MM force fields are not sufficiently accurate as to reproduce ab initio quality electrostatic potentials. Overcoming this problem requires extending MM force fields with an explicit description of polarization, an open problem that has been the subject of intense research over the past decade.[1−25,67−70]

Significant effort has been focused on the development of both polarizable protein force fields[6−15,67−70] and polariz-

* Corresponding author e-mail: victor.batista@yale.edu.
[†] Yale University.
[‡] Los Alamos National Laboratory.

10.1021/ct050218h CCC: $33.50
Published on Web 11/18/2005

able models for small molecules.[15−25] While these methods are expected to become routine practice, no polarizable force field has so far been widely implemented for protein modeling. Parameters are still under development, and published applications are limited to those from the development groups. This is partly due to the inherent difficulty of the polarization problem and the fact that the behavior of polarizable force fields for flexible molecules (e.g., amino acids) has yet to be fully understood.[24] Also, the methods and software required to treat polarization are not as standardized as for the pairwise protein potentials. Finally, the increased complexity and expense of polarizable force fields make their applications to protein modeling justifiable only when introducing significant corrections.

Semiempirical QM approaches, based on linear-scaling methods, are nowadays capable of calculating molecular electrostatic potentials for systems as large as proteins.[71−74] Comparisons to benchmark calculations, however, indicate that accurate calculations of electrostatic potentials would still require the development of more reliable semiempirical methods,[75−77] a problem that remains a subject of much current research interest.[78−81]

Considering the central role of electrostatic interactions in biological systems, it is therefore imperative to develop accurate, yet practical, approaches for describing molecular electrostatic potentials. To this end, the first objective is the development of a computational protocol capable of providing accurate electrostatic potentials for proteins in well-defined configurations. The protocol introduced in this paper addresses such a computational task by computing protein electrostatic potentials according to rigorous ab initio quantum chemistry methods. Under the new protocol, the protein is partitioned into molecular domains according to a simple space-domain decomposition scheme. ElectroStatic-Potential (ESP) atomic charges of the constituent domains are iteratively computed until reaching convergence in the description of the protein electrostatic potential. Such an iterative scheme scales linearly with the size the system, bypassing the enormous demands of memory and computational resources that would be required by a brute-force quantum chemistry calculation of the complete system. The accuracy and capabilities of the method are demonstrated in applications to benchmark calculations as well as in studies of the electrostatic potential in the potassium ion channel and electrostatic contributions to protein−protein interactions.

The paper is organized as follows. Section 2.1 describes the specific QM/MM methodology applied in this study. Section 2.2 describes the space-domain decomposition scheme for computations of electrostatic potentials. The computational details regarding the calculation of ESP charges are outlined in the Appendix. Results are presented in section 3, including applications to calculations of electrostatic potentials in the potassium channel protein and the description of protein−protein electrostatic interactions in the barnase-barstar complex, modeling solvation effects according to the Poisson−Boltzmann equation. Section 4 summarizes and concludes.

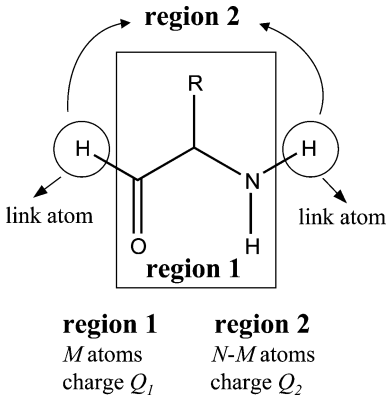


Figure 1. Representation of the regions used in the fitting procedure for each QM/MM calculation. Although a single residue is shown, region 1 may actually contain more than one residue.

2. Methods

2.1. QM/MM Methodology. The computational protocol, detailed in section 2.2, can be implemented in combination with any QM/MM hybrid method in which the polarization of the QM region due to the electrostatic influence of the surrounding molecular environment is explicitly considered. The particular QM/MM methodology applied in this study is the ONIOM-EE (HF/6-31G*:Amber) approach,[82−88] as implemented in Gaussian03,[89] with QM and MM layers defined in Figure 1.[90]

In the ONIOM-EE approach, the molecular domain of interest (herein called region *X*) is treated according to rigorous ab initio quantum chemistry methods, while the rest of the system (herein called region *Y*) is treated according to MM force fields. For systems where regions *X* and *Y* are covalently bonded, a QM/MM boundary is defined, and the covalency of frontier atoms is completed according to the standard link-hydrogen atom scheme.

The computation of a molecular property *A* (e.g., the energy, or the molecular electrostatic potential) involves the combination of three independent calculations:

$$A = A_{X+Y}(\mathrm{MM}) + A_X(\mathrm{QM}) - A_X(\mathrm{MM}) \qquad (1)$$

Here, A_{X+Y}(MM) is the property of interest, modeled at the MM level of theory for the complete system, while A_X(QM) and A_X(MM) are the same property of the reduced-system computed at the QM and MM levels of theory, respectively.

The effect of electrostatic interactions between the QM and MM layers is included in the calculation of both A_X(QM) and A_X(MM). In particular, A_X(QM) includes the effect of electrostatic interactions between the distribution of charges in the MM region and the electronic density of the QM layer obtained according to ab initio quantum chemistry methods. In addition, the contributions due to electrostatic interactions between regions *X* and *Y*, modeled at the MM level, are included in the calculation of both A_X(MM) and A_{X+Y}(MM) and therefore cancel out. The resulting evaluation of molecular properties thus includes a QM description of polarization of the reduced system, as influenced by the surrounding protein environment, while van der Waals interactions between *X* and *Y* are described at the MM level. For

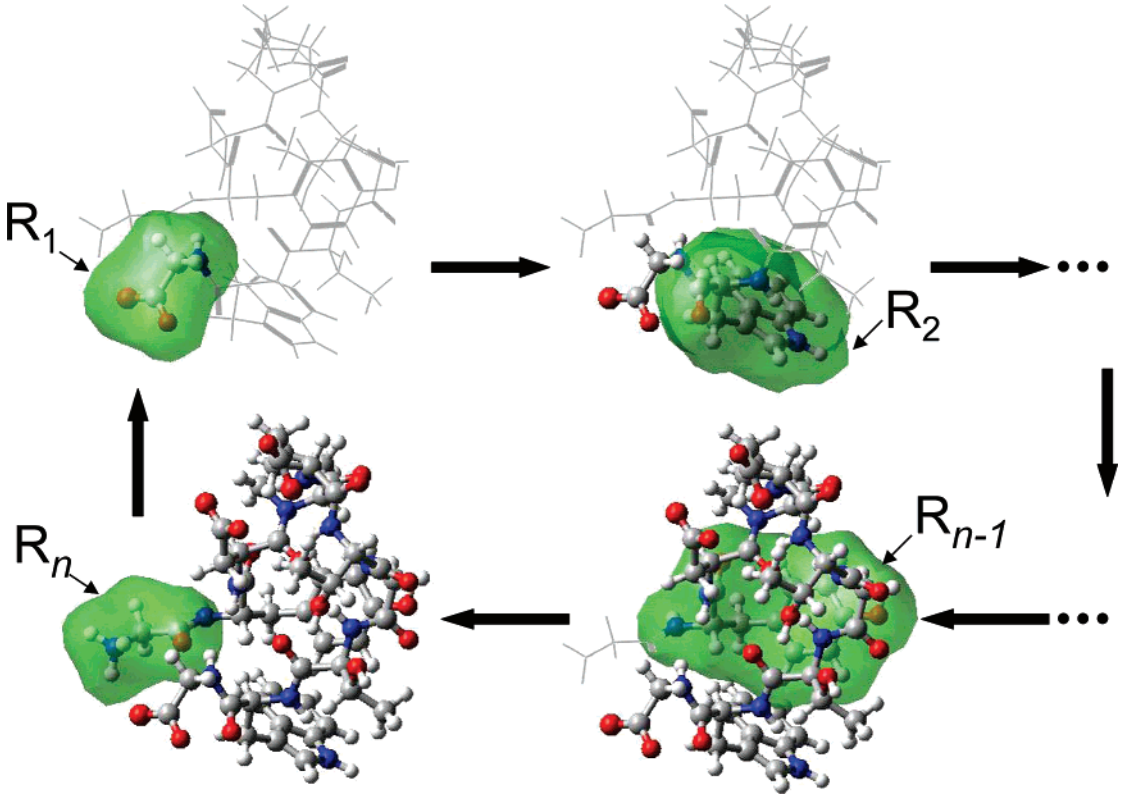


Figure 2. Representation of the MoD-QM/MM method. Green surfaces represent the QM region in QM/MM calculations. Colored balls and sticks represent regions with updated charges.

comparison, QM/MM calculations where the QM layer is *not* polarized by the surrounding environment are performed at the *ONIOM Molecular-Embedding* (ONIOM-ME) level of theory. In this QM/MM approach A_X(QM) and A_X(MM) do *not* include electrostatic interactions between regions *X* and *Y*.

Determination of ESP atomic charges is based on a least-squares minimization procedure where the electrostatic potential due to the ESP charges is fitted to the QM/MM electrostatic potential computed over a set of grid points around the QM layer. A detailed description of the calculation of ESP atomic charges, subject to the boundary conditions imposed by the link-hydrogen atom scheme, is presented in the Appendix.

2.2. Space-Domain Decomposition Scheme. Consider the task of modeling the molecular electrostatic potential of a polypeptide in a well-defined configuration (e.g., the X-ray structure). For a small polypeptide, such a calculation can be accomplished by first computing the molecular electronic density, according to rigorous ab initio quantum chemistry methods, and subsequently fitting the electrostatic potential on a set of grid points around the molecule to a standard multipole expansion.[78,91−93]

The simplest model truncates the multipole expansion after the monopole term, thus requiring only the calculation of ESP atomic charges. While rigorous, such a calculation is computationally intractable for large systems (e.g., proteins) due to the overwhelming demands of memory and computational resources that would be required by 'brute-force' quantum chemistry calculations of the complete system. As a result, it is common practice to approximate protein electrostatic potentials as a sum of the electrostatic potentials of the constituent molecular fragments (e.g., amino acid residues), neglecting the mutual polarization effects. Computations based on popular MM force fields[61−66] as well as studies of protein docking[42,48] or activity relationships[94,95] are based on such an approximation, even though breakdown of this assumption is the rule rather than the exception whenever there are charged or polar fragments (e.g., amino acid residues) in the system. It would, therefore, prove a significant advance to extend such a methodology to compute distributions of ESP atomic charges where polarization effects are explicitly considered.

Motivated by the necessity to avoid a 'brute-force' quantum chemistry calculation of the complete system, an iterative space-domain decomposition scheme is introduced (see Figure 2): the system is partitioned into molecular domains (green regions in Figure 2) of suitable size for efficient quantum chemistry calculations. For simplicity, proteins are partitioned into *n* molecular domains containing amino acid residues R_1, R_2, ..., R_n, although more general partitioning schemes could be considered analogously (e.g., partitions containing more than one residue, ions, and solvent molecules). The computation of the protein electrostatic potential can then be accomplished as follows. Starting with a QM layer containing amino acid residue R_1 (see top-left panel of Figure 2, green region), the ESP atomic charges of R_1 are computed according to the QM/MM hybrid methods that explicitly consider the electrostatic influence of the MM layer describing the surrounding protein environment. Next, the QM layer is redefined as a molecular domain containing amino acid residue R_2 (see top-right panel of Figure 2, green

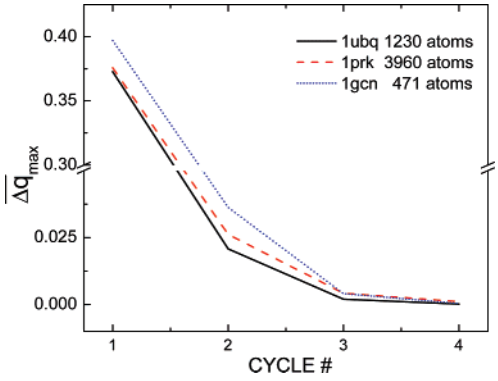


Figure 3. Maximum atomic charge difference (in atomic units) between successive iterations, averaged over all residues, as a function of the MoD-QM/MM iteration cycle. Three representative proteins are shown, including ubiquitin from *Human erythrocytes* (1ubq), proteinase K (2prk) from *Tritirachium album*, and glucagon (1gcn) from *Sus scrofa*.

region). The atomic charges of R_1, now in the MM layer (balls and sticks, top-right panel of Figure 2), are updated according to the ESP charges obtained in the previous step. The ESP atomic charges of amino acid residue R_2 are computed analogously, and the procedure is subsequently applied to the remaining set of molecular domains containing amino acid residues R_3 ... R_n. Note that each calculation of atomic charges considers the updated distribution of charges on all previously considered molecular domains. The entire computational cycle is subsequently iterated several times until reaching self-consistency.

The resulting methodology (called 'Moving Domain-QM/MM' (MoD-QM/MM) approach throughout this manuscript) converges within a few iteration cycles (i.e., usually 4 or 5 cycles), scaling linearly with the size of the system (i.e., the total computational time is $\tau = N_c \times \tau_0 \times n$, where $N_c \approx 4$ is the number of iteration cycles needed for convergence, τ_0 is the average computational time required for a single-point calculation of an individual molecular domain, typically a few minutes, and n is the number of molecular domains in the protein). The advantage of the resulting electrostatic potential, relative to other models based on static point-charge model distributions,[61−66] is that the MoD-QM/MM approach explicitly considers mutual polarization effects between amino acid residues, providing ab initio quality electrostatic potentials (see section 3). The accuracy of the resulting molecular electrostatic potential, however, comes at the expense of *transferability* since the computed distribution of atomic charges is in principle *nontransferable* to other protein configurations. Therefore, while accurate, the computed electrostatic potential is useful only for applications where conformational changes are negligible.

Figure 3 illustrates typical convergence rates for the implementation of the MoD-QM/MM computational protocol, as applied to the calculation of the molecular electrostatic potentials of three representative protein structures downloaded from the Protein Data Bank, including Ubiquitin from *Human erythrocytes* (1ubq), solved at 1.8 Å resolution,[96] Proteinase K (2prk) from *Tritirachium album*, solved at 1.5 Å resolution,[97] and Glucagon (1gcn) from *Sus scrofa*, solved at 3.0 Å resolution.[98] Figure 3 shows a convergence measure

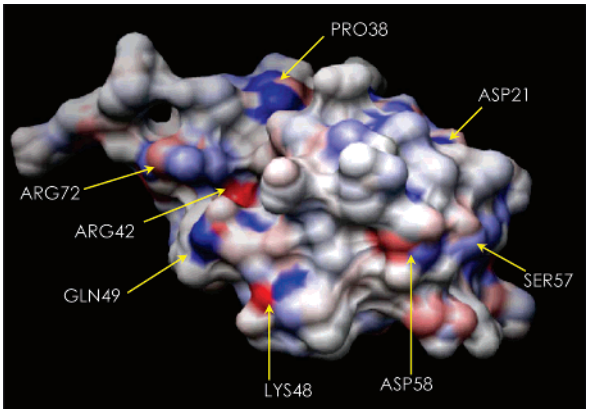


Figure 4. Surface of ubiquitin colored according to the differences in atomic charges obtained by considering, or neglecting, the mutual electrostatic influence between amino acid residues at the ONIOM-EE and ONIOM-ME levels of theory, respectively. Blue(red) color indicates an increase-(decrease) in electronic density due to polarization effects (maximum differences, indicated by bright coloring, correspond to changes of atomic charges of the order of ±20%).

as a function of the iteration cycle, defined as the maximum change in atomic charges per residue averaged over all amino acid residues in the protein. It is shown that self-consistency is typically achieved within four iteration cycles, a convergence rate that is found to be independent of the system size. It is also found that the convergence rate is independent of the order chosen for treating individual molecular domains in each cycle.

To illustrate typical results of protein polarization, as modeled by the MoD-QM/MM protocol, Figure 4 shows a color map of the 1ubq surface displaying differences in atomic charges obtained by considering, or neglecting, the mutual electrostatic influence between amino acid residues at the ONIOM-EE and ONIOM-ME levels of theory, respectively. It is shown that typical corrections to atomic charges of specific amino acid residues can be as large as 20% due to polarization effects. These corrections are thus expected to be important in applications where there is collective electrostatic influence with contributions from several residues.

3. Results

Results are presented in three subsections. Section 3.1 demonstrates the capabilities of the MoD-QM/MM methodology for reproducing ab initio electrostatic potentials associated with the so-called 'molecular bottleneck' in the potassium channel protein from *streptomyces lividans*.[99] Section 3.2 implements the MoD-QM/MM method, in conjunction with the Poisson−Boltzmann equation, in applications to the description of protein−protein electrostatic interactions. Finally, section 3.3 analyzes the capabilities of the MoD-QM/MM method for generating a data bank of electrostatic potentials associated with several proteins in their X-ray structure configurations.

3.1. Potassium Ion Channel. This section illustrates the implementation of the MoD-QM/MM approach as applied to the description of the molecular electrostatic potential of

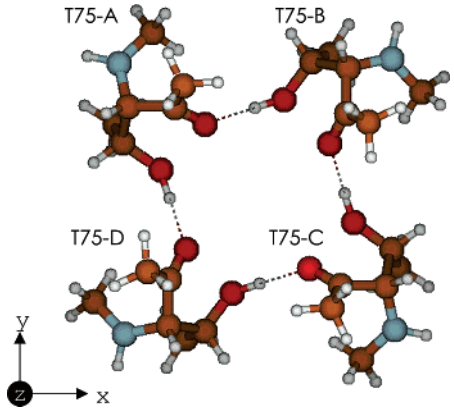


Figure 5. Structure of the complex formed by threonine residues (THR-75) of the four identical subunits forming the selectivity filter in the KscA potassium channel.

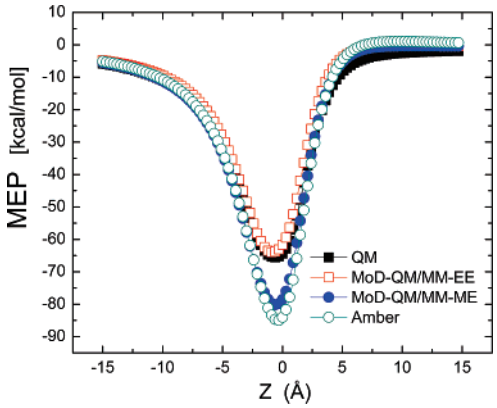


Figure 6. Molecular electrostatic potential (MEP) along the central axis of the tetramer of threonine residues (THR-75) in the KscA potassium channel.[99] The MEP is calculated according to four different methods: full Quantum Mechanics at the HF/6-31G* level of theory (open square); atomic charges obtained with the MoD-QM/MM-EE approach (solid square); MoD-QM/MM at the ONIOM-ME (HF/6-31G*:Amber) level (i.e: neglecting polarization) (solid circle); and Amber MM force field charges (open circle).

the potassium channel protein from *Streptomyces lividans* (KcsA K^+ channel),[39,99] with emphasis on benchmark calculations on truncated and QM/MM models of the so-called *selectivity filter*.

System (**1**) involves a truncated tetramer benchmark model amenable to rigorous ab initio quantum chemistry calculations (see Figure 5). The model involves 88 atoms and includes only the residues THR-75 belonging to the four identical peptide chains that constitute the ion channel. Such a tetramer provides the largest contribution to thes molecular electrostatic potential at the selectivity filter. The structural model is built according to the configuration of the THR-75 tetramer in the X-ray crystal structure of the KcsA K^+ channel (PDB access code 1bl8), adding hydrogen atoms and capping both ends of the THR residues with methyl groups. The MoD-QM/MM approach is implemented by partitioning the tetramer into four molecular domains defined by the individual THR-75 residues capped with methyl groups.

Figure 6 compares calculations of the electrostatic potential evaluated along the central axis of the ion channel (see Figure 5, *z*-axis) according to four different methodologies, including the ab initio HF/6-31G* level, the MoD-QM/MM approach implemented at the ONIOM-EE (HF/6−31G*:Amber) level of theory (MoD-QM/MM-EE), and methods where polarization effects are neglected, including both the Amber MM force field and the MoD-QM/MM approach implemented at the ONIOM-ME (HF/6-31G*:Amber) level of theory (MoD-QM/MM-ME). The molecular electrostatic potential at position *z* is expressed in kcal/mol as the interaction energy felt by a unit of positive charge at position *z*. Figure 6 shows that the MoD-QM/MM-EE results are in excellent agreement with benchmark ab initio calculations. In contrast, calculations where polarization effects are neglected deviate strongly, overestimating the molecular electrostatic potential by more than 20 kcal/mol. In particular, the electrostatic potential obtained at the ONIOM-ME (HF/6-31G*:Amber) level of theory (i.e., neglecting mutual polarization effects between the four separate THR residues) is in very good agreement with the description provided by the Amber MM force field. These results indicate that deviations between ab initio and MM results are mainly due to the intrinsic approximation of MM force fields, based on transferable static point-charge model distributions that neglect polarization effects. Furthermore, the agreement between ab initio and MoD-QM/MM-EE calculations indicates that such polarization effects can be quantitatively addressed by the static point-charge model distributions generated according to the MoD-QM/MM-EE method, providing ab initio quality electrostatic potentials.

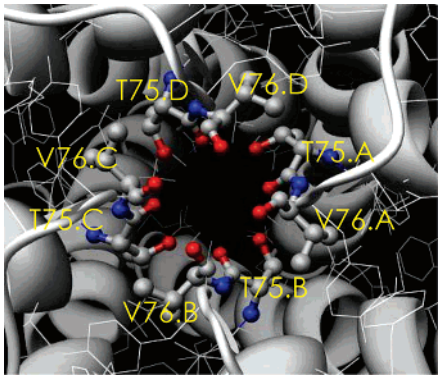


Figure 7. Structure of the complex formed by THR-75 and VAL-76 residues embedded in the KscA potassium channel.

System (**2**) involves a QM/MM structural model of the complete KcsA K^+ channel with an expanded QM layer of 128 atoms that includes both THR-75 and VAL-76 residues of the four identical polypeptide subunits forming the selectivity filter (see Figure 7). The rest of the protein is treated at the MM level. The model allows one to address the capabilities of the MoD-QM/MM computational protocol as applied to the description of polarization of an extended QM layer due to the influence of the surrounding protein environment.

The structural model of the entire protein is prepared according to the X-ray crystal configuration of the KcsA K^+ channel (PDB access code 1bl8) adding hydrogens and partially relaxing the protein configuration, keeping α-carbons fixed at their crystallographic positions in order to

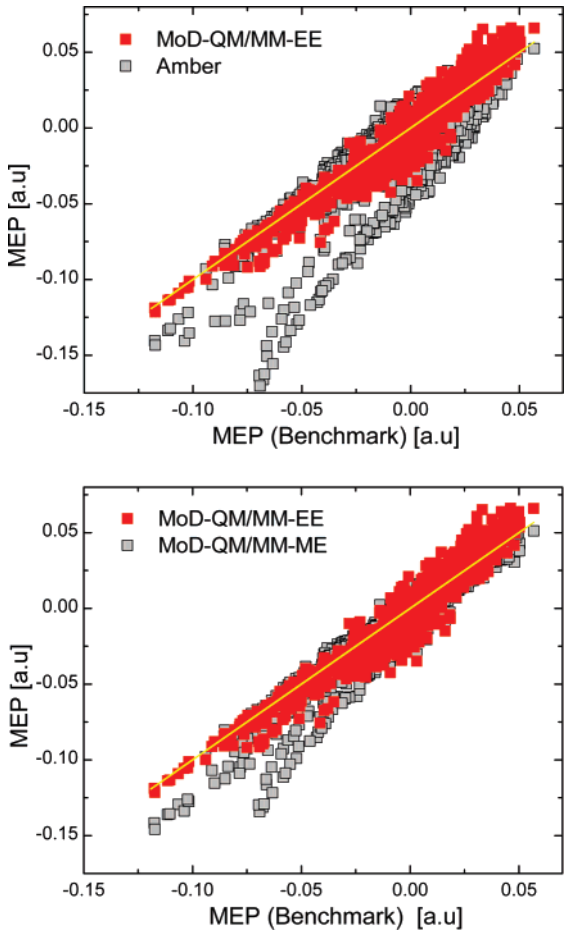


Figure 8. Correlation between the Molecular Electrostatic Potentials (MEP) (in atomic units) obtained according to the MoD-QM/MM-EE approach (red points) and benchmark QM/MM calculations for a distribution of grid of points around the tetramer of THR-75 and VAL-76 residues. The yellow line indicates complete correlation with benchmark calculations. The upper and lower panels compare the correlation of MEP obtained according to the MoD-QM/MM-EE approach (red points) with the corresponding results obtained by neglecting polarization effects according to the Amber MM force field (upper panel, gray points) and the MoD-QM/MM-ME method (lower panel, gray points).

preserve the natural shape of the protein. Benchmark calculations of the molecular electrostatic potential, computed at the ONIOM-EE (HF/6-31G*:Amber) level of theory, are compared to the corresponding results obtained according to the MoD-QM/MM-EE method, the Amber MM force field, and the MoD-QM/MM-ME approach. The MoD-QM/MM methodology is implemented by partitioning the system into four molecular domains. Each domain includes a THR-75/VAL-76 pair of residues with link hydrogen atoms placed at the amide bonds between THR-74 and THR-75 and between VAL-76 and GLY-77.

Figure 8 shows the correlation between MoD-QM/MM and benchmark QM/MM calculations of the electrostatic potential evaluated on a distribution of grid points away from the *z* axis of the potassium channel. The grid involves a lattice of 3000 points distributed in three layers at 2.5 grid points/Å around the extended tetramer of THR-75 and VAL-76 residues. For completeness, Figure 8 also includes the analysis of correlations between benchmark ab initio calculations and the corresponding results obtained according to the Amber MM force field (upper panel, gray points) and the MoD-QM/MM-ME approach (see Figure 8, lower panel, gray points) where polarization effects are neglected.

Deviations relative to complete correlation are quantified over the set of N_g grid points in terms of the root-mean squared deviation

$$\xi = [\sum_{i=1}^{N_g} (u_i - U_i)^2/N_g]^{1/2} \quad (2)$$

Here, U_i is the reference QM/MM electrostatic potential evaluated at grid point i and u_i is the electrostatic potential generated according to the static point-charge models generated by the MoD-QM/MM-EE, MoD-QM/MM-ME methods or the Amber MM force field. Root-mean-squared deviations $\xi = 4.9$, 6.6, and 11.3 kcal/mol/C are obtained when using the MoD-QM/MM-EE, MoD-QM/MM-ME, and Amber MM force field methods, respectively. These results indicate that the MoD-QM/MM-EE approach correlates significantly better with benchmark calculations than methods where polarization effects are neglected.

3.2. Protein−Protein Interactions. The binding energy of protein−protein complexes often depends on a delicate balance of several factors, including hydrophobic and electrostatic energy contributions associated with protein−protein and solvent−protein interactions.[100] Computations based on continuum electrostatic methods[101] suggest that even complementary Coulombic interactions that stabilize protein−protein complexes are usually not strong enough to compensate for unfavorable desolvation effects.[102−107] Therefore, the driving force for complexation is generally expected to come mainly from nonpolar interactions.[106,108] However, continuum electrostatic calculations are usually based on inaccurate molecular electrostatic potentials provided by nonpolarizable MM force fields. Therefore, it is natural to expect that calculations based on more accurate electrostatic potentials might provide further insight on the role played by electrostatic interactions in the process of protein−protein complexation.

This section applies the MoD-QM/MM approach in conjunction with the methods of continuum electrostatics in order to analyze the electrostatic contributions to the binding energy of the complex formed by the extracellular ribonuclease barnase and its intracellular inhibitor, the protein barstar. Such a complex system is ideally suited to investigate the capabilities of the MoD-QM/MM approach for explicitly modeling polarization effects because the complex has been extensively investigated both theoretically and experimentally.[49,109−113]

The barnase−barstar complex involves complementary proteins that bind fast and with high affinity. The binding interface involves mainly polar and charged residues as well as several bound-water molecules stabilizing the complex through complementary electrostatic interactions. However, the desolvation energy of charged and polar residues destabilizes the complex. Previous theoretical studies, based on continuum solvent models,[49,109−113] show contradictory

results regarding the analysis of stabilizing and destabilizing factors. Studies include reports of an unfavorable electrostatic binding energy of +14 kcal/mol,[111] a near zero electrostatic contribution to the binding energy (with desolvation and complexation terms almost canceling each other),[49] and finally, a favorable electrostatic contribution when considering a high protein dielectric constant.[112] The main objective of this section is to address this controversial aspect of the problem, recalculating the electrostatic contributions to the binding energy of the barnase−barstar complex according to the same methods of continuum electrostatics, using a distribution of atomic charges of the complex obtained according to the MoD-QM/MM-EE approach.

The structure of the barnase−barstar complex is prepared according to ref 49. The electrostatic contribution to the binding energy of complexation of barnase (A) and barstar (B) to form the barnase−barstar complex (AB) is defined as

$$\Delta\Delta G_{\text{elec}} = \Delta G_{\text{elec}}(AB) - \Delta G_{\text{elec}}(A) - \Delta G_{\text{elec}}(B) \quad (3)$$

where $\Delta G_{\text{elec}}(\xi)$ represents the electrostatic free-energy of the macromolecular system ξ

$$\Delta G_{\text{elec}}(\xi) = \frac{1}{2}\sum_i q_i \phi(\mathbf{r}_i) \quad (4)$$

where ξ is either *A*, *B*, or *AB* and the summation is carried out over all atomic charges q_i in ξ.

The electrostatic potential $\phi(\mathbf{r}_i)$, corresponding to charges q_i placed at $\mathbf{r}_i$, is obtained by solving the finite-difference Poisson−Boltzmann equation[114,115] with Delphi.[116] The interiors of the protein complex and aqueous solution are modeled as continuum media with dielectric constants $\epsilon_p = 2$ and $\epsilon_w = 80$, respectively. The choice of $\epsilon_p = 2$ for the dielectric constant of the protein interior is consistent with previous studies based on the assumption that complexation does not involve conformational changes but only electronic relaxation.[49] Boundary conditions are approximated by the Debye-Hückel potential of the charge distribution. The total energy calculations is converged within 10^{-4} k_BT, where k_B is the Boltzmann constant and T is the absolute room-temperature. Atomic radii are defined according to the CHARMM MM force field.[66]

The electrostatic contributions to the free-energy of complexation $\Delta\Delta G_{\text{elec}}$ is −12.6 kcal/mol, when using the distribution of atomic charges given by the MoD-QM/MM-EE protocol, with 0.1 M ionic strength of the aqueous solution and 1.4 Å for the ionic exclusion radius, indicating significant electrostatic stabilization of the complex. In contrast, the electrostatic contributions computed by using the CHARMM distribution of atomic charges, where protein polarization effects are not explicitly considered, is $\Delta\Delta G_{\text{elec}} = 3.3$ kcal/mol, in agreement with previous calculations.[49] These results indicate that the overall electrostatic stabilization of the complex is mainly due to protein polarization over the extended protein−protein contact surface.

It has been recognized that the results of Poisson−Boltzmann calculations depend rather sensitively on the atomic radii. In fact, a set of atomic Born radii has been obtained by Roux and co-workers[117] to reproduce quantitatively the electrostatic contributions to the solvation free energy of the 20 natural amino acids, computed by free energy perturbation techniques, performing Poisson-Boltzmann calculations with the CHARMM MM force field. Using such a set of atomic radii we obtain $\Delta\Delta G_{\text{elec}} = -3.3$ kcal/mol for 0.1 M ionic strength of the aqueous solution, when using the atomic charges prescribed by the CHARMM MM force field and $\Delta\Delta G_{\text{elec}} = -23.0$ kcal/mol when using the atomic charges obtained according to the MoD-QM/MM-EE protocol in close agreement with the experimental value $\Delta\Delta G_{\text{elec}} = -19.0$ kcal/mol.[118]

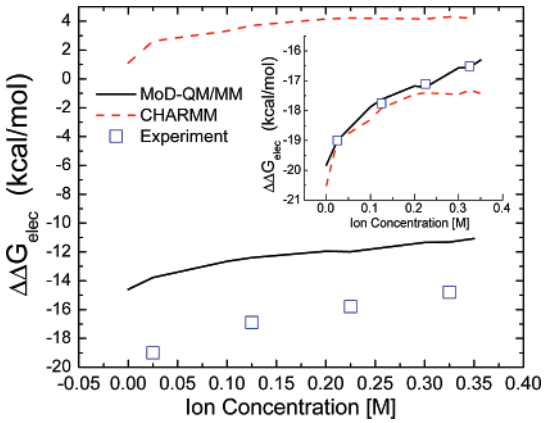


Figure 9. Calculated and experimental[118] binding energies as a function of the ionic strength. The inset shows all curves artificially shifted to make them coincide at 25 mM, to facilitate the comparison of ionic strength dependency.

For completeness, Figure 9 compares experimental binding energies[118] as a function of ionic strength and the corresponding electrostatic contributions to the binding energy computed by using the distribution of atomic charges provided by the MoD-QM/MM approach and the CHARMM MM force field. These results indicate that electrostatic interactions, as described by the MoD-QM/MM protocol, play a dominant role in the overall stabilization of protein complexes and reproduce the experimental dependence of the binding stability as a function of the solution ionic strength.

The observation that polarization effects play a dominant role in the overall stabilization of the complex barnase−barstar leads to the following questions: What residues are more significantly polarized? What are the specific interactions responsible for polarization of individual residues? To address these questions, a detailed analysis of electrostatic contributions is performed. The binding energy of the complex is recomputed, after substituting the polarized charges of individual residues obtained at the ONIOM-EE level by the unpolarized charges obtained at ONIOM-ME level of theory. The electrostatic contribution to the total binding energy of the complex, due to polarization of residue *i*, is then defined as the resulting change in binding energy $\Delta\Delta G^i_{\text{elec}}$.

The upper and lower panels of Figure 10 show the results of $\Delta\Delta G^i_{\text{elec}}$ for all residues in barnase and barstar, respectively. It is shown that the largest contribution to the binding

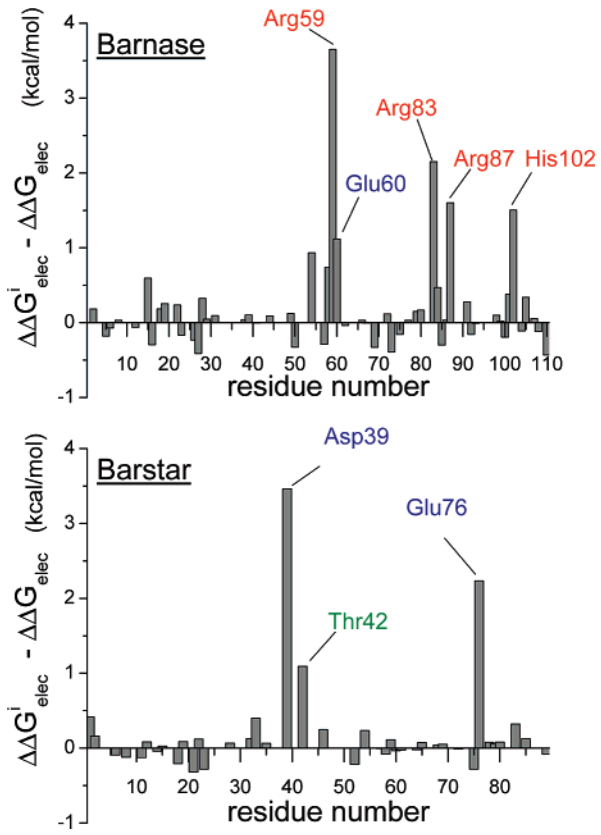


Figure 10. $\Delta\Delta G^i_{elec}$ in the upper and lower panels represents the binding energy of the complex after replacing the MoD-QM/MM-EE charges on residue *i* by MoD-QM/MM-ME charges. $\Delta\Delta G_{elec}$ is the binding energy as reported in the text (i.e. using MoD-QM/MM-EE charges for all residues).

energy due to polarization of individual residues in barnase results from the amino acid residues Arg-59, Arg-83, and Arg-87, while residues Asp-39 and Glu-76 provide the most important contributions in barstar. Not surprisingly, all of these residues are located at the complex interface and polarize each other through specific residue−residue interactions. In particular, Arg-83 and Arg-87 are polarized by Asp-39. Similarly Arg-59 forms a salt bridge with Glu-76. These results strongly suggest that the specific polarization of multiple pairs of amino acid residues at the barnase−bastar interface is largely responsible for the binding energy of the complex.

3.3. MM Force Fields. The calculations reported in previous sections illustrate the well-known fact that nonpolarizable MM force fields (i.e., based on transferable static point-charge distributions) provide only approximate descriptions of molecular electrostatic potentials, commonly exhibiting significant deviations from benchmark ab initio calculations. In contrast, the static point-charge model distributions generated according to the MoD-QM/MM protocol are capable of providing more accurate electrostatic models, at least when the systems remain near the reference (e.g., X-ray structure) configurations. Considering that there is a wide range of applications where conformational changes can be neglected, it is important to consider whether the MoD-QM/MM protocol can be applied to generate a data bank of ab initio quality electrostatic potentials based on static point-charge model distributions, with emphasis on proteins at reference (e.g., X-ray structure) configurations. Furthermore, it is important to analyze whether such polarized static point-charge model distributions can be used to reparametrize standard MM force fields in an effort to improve their description of electrostatic potentials of specific proteins near their corresponding reference configurations.

Table 1. Root Mean Square Deviation (RMSD) of Optimized Protein Structures Relative to the Crystal Structure[a,b]

protein	Amber	MoD-QM/MM-EE	OPLS	PFF
1MAG	1.71	1.37		
1PI8	1.76	1.44		
1UBQ	1.97	1.85	2.08	1.97
2PRK	0.99	0.93	1.26	1.33
1PGX	2.32	1.97	4.1	4.02
1GCR	1.55	1.32	1.53	1.54
1GCN	5.08	2.12	4.14	3.77
1SSI	1.64	1.54	1.93	1.89
2RN2	2.11	1.79	1.92	1.54
1LTD	1.41	1.00		
average	2.06	1.53	2.43	2.29

[a] Reference 119. [b] RMSD values are reported in Å for molecular structures obtained by using the Amber MM force field, the MoD-QM/MM-EE method, the Polarizable Force Field (PFF),[119] and the OPLS MM force field.[65]

Reparametrization of the Amber MM force field according to the distribution of atomic charges generated by the MoD-QM/MM protocol would, in principle, requires a subsequent readjustment of the torsional coefficients.[64] In practice, however, torsional parameters are expected to remain almost unchanged so long as the minimum energy configuration is sufficiently similar to the reference (e.g., X-ray) structure. It is, therefore, expected that an approximate MM force field constructed by substituting the Amber charges by the atomic charges generated according to the MoD-QM/MM protocol could be sufficiently accurate as to provide a reliable description of both electrostatic and steric interactions whenever the system remains near the reference configuration.

To investigate the effect of charge reparametrization as applied to the Amber MM force field, 10 realistic protein structures from the Protein Data Bank (listed in Table 1) were used as initial geometries for gas-phase energy minimization after substituting the original RESP charges by ESP atomic charges generated according to the MoD-QM/MM protocol.

Table 1 shows the root-mean-square deviations (RMSD) relative to reference X-ray structures. For comparison, results obtained with four different approaches are shown, including the Amber MM force field, as parametrized with RESP charges; the Amber MM force field with atomic charges computed according to the MoD-QM/MM-EE protocol; the OPLS-AA MM force field;[65] and finally, results obtained with a Polarizable Force field (PFF).[119] In all cases, geometry minimization procedures were performed using a convergence criterion of 0.05 kcal/mol/Å for the root-mean-square gradient. Since these proteins were resolved at a high resolution, it is reasonable to expect a low RMSD to be an

indication of how well the resulting force field describes the corresponding reference configurations. Table 1 shows that the RMSD obtained by using the Amber MM force field with MoD-QM/MM-EE atomic charges favorably compares to other alternative approaches. While still approximate, the resulting modified force field is thus expected to provide not only better quality electrostatic potentials than those provided by the original MM force field but also minimum energy configurations more similar to the reference X-ray crystal structures.

4. Conclusions

We have introduced the MoD-QM/MM computational protocol to account for protein polarization effects when computing molecular electrostatic potentials according to static point-charge model distributions. The method implements an iterative space-domain decomposition scheme, partitioning the protein into molecular domains of suitable size for efficient quantum chemistry calculations. ESP atomic charges are then computed, in a self-consistent manner, according to QM/MM hybrid methods that explicitly include polarization effects due to the electrostatic influence of the surrounding protein environment. The resulting methodology usually converges within a few iteration cycles, regardless of the protein size. Therefore, the overall computational cost scales *linearly* with the size of the system, bypassing the enormous demands of computational resources that would be required by brute-force quantum chemistry calculations of the complete protein.

We have shown that quantitative agreement with ab initio calculations is verified in the description of electrostatic potentials of small polypeptides benchmark systems where polarization effects are significant, showing a remarkable improvement relative to the corresponding electrostatic potentials obtained with popular MM force fields. Furthermore, the application of the MoD-QM/MM method to the QM/MM description of the potassium channel of *streptomyces lividans* demonstrates the capabilities of the protocol for modeling polarization effects induced by the surrounding protein environment on the selectivity filter.

We showed that the MoD-QM/MM protocol, implemented in conjunction with methods of continuum electrostatics, offers a particularly promising methodology for studies of protein−protein interactions where protein polarization effects are explicitly considered. The application of such a combined methodology to calculations of electrostatic contributions to the binding energy of the barnase−barstar complex indicates that polarization of the protein−protein interface can lead to significant electrostatic stabilization of the complex. Furthermore, we have shown that such an electrostatic contribution is most responsible for the overall dependency of the total binding free-energy with ionic strength.

We have demonstrated the feasibility of constructing a data bank of electrostatic potentials based on static point-charge model distributions corresponding to protein structures from the Protein Data Bank. Finally, we have implemented the generated electrostatic potentials in conjunction with the Amber MM force field in an effort to improve the description of electrostatic potentials provided by MM force fields and generate relaxed minimum energy configurations more similar to reference high-resolution X-ray crystal structures.

Acknowledgment. V.S.B. acknowledges a generous allocation of supercomputer time from the National Energy Research Scientific Computing (NERSC) center and financial support from Research Corporation, Research Innovation Award # RI0702, a Petroleum Research Fund Award from the American Chemical Society PRF # 37789- G6, a junior faculty award from the F. Warren Hellman Family, the National Science Foundation (NSF) Career Program Award CHE # 0345984, the NSF Nanoscale Exploratory Research (NER) Award ECS # 0404191, the Alfred P. Sloan Fellowship (2005−2006), a Camille Dreyfus Teacher-Scholar Award for 2005, and a Yale Junior Faculty Fellowship in the Natural Sciences (2005−2006). S.S.F.L. acknowledges a summer research fellowship and hospitality at LANL and support from the NIH predoctoral training grant T32 GM008283. S.S.F.L and E.R.B acknowledge funding from the LDRD program at LANL and the Heavy element chemistry BES-DOE program. The authors are grateful to Mr. Sabas Abuabara for proofreading the manuscript.

Appendix: QM/MM Computation of ESP Atomic Charges

This section describes the implementation of boundary conditions imposed by the link-hydrogen atom scheme for computations of ESP atomic charges of an individual amino acid residue, as polarized by the surrounding protein environment.

For a given QM/MM calculation, N is the total number of atoms in the QM layer, including M atoms within the residue and $N - M$ link atoms (see Figure 1). The total charge of the QM layer is

$$Q = Q_1 + Q_2 \tag{A-1}$$

where

$$Q_1 = \sum_{i=1}^{M-1} q_i + q_M,\ Q_2 = \sum_{i=M+1}^{N-1} q_i + q_N \tag{A-2}$$

where Q_1 is the net charge of the residue and Q_2 is set equal to zero in order to ensure consistency with standard MM force fields.

The electrostatic potential at position $\mathbf{r}_j$ due to all point charges in the QM region is written as

$$u_j = \sum_{i=1}^{N} \frac{q_i}{r_{ji}} \tag{A-3}$$

where $r_{ji} \equiv |\mathbf{r}_j - \mathbf{r}_i|$. Since we impose conditions (A-2) for the charge in regions 1 and 2, eq A-3 can be written as follows:

$$u_j = \sum_{i=1}^{M-1} \left[\frac{q_i}{r_{ji}} - \frac{q_i}{r_{jM}}\right] + \frac{Q_1}{r_{jM}} + \sum_{i=M+1}^{N-1} \left[\frac{q_i}{r_{ji}} - \frac{q_i}{r_{jN}}\right] + \frac{Q_2}{r_{jN}} \tag{A-4}$$

Making the substitutions, $F_{jik} \equiv (1/r_{ji} - 1/r_{jk})$ and $K_{jk} \equiv 1/r_{jk}$, eq A-4 can be rewritten as follows:

$$u_j = \sum_{i=1}^{M-1} q_i F_{jiM} + \sum_{i=M+1}^{N-1} q_i F_{jiN} + Q_1 K_{jM} + Q_2 K_{jN} \quad \text{(A-5)}$$

The actual computation of ESP atomic charges q_i requires a least-squares minimization of the χ^2 error function

$$\chi^2 = \sum_{j}^{N_g} (u_j - U_j)^2 \quad \text{(A-6)}$$

where U_j is the QM/MM electrostatic potential at grid point j and u_j is the corresponding electrostatic potential defined by the distribution of point charges. The summation, introduced by eq A-6, is carried over a set of N_g grid points, associated with four layers of grid points at 1.4, 1.6, 1.8, and 2.0 times the van der Waals radii around the QM region, each of them with a density of 1 grid point Å^{-2}.

From eq A-6, the minimum of χ^2 can be obtained by imposing the condition

$$\frac{\partial \chi^2}{\partial q_k} = -\sum_{j=1}^{N_g} 2(u_j - U_j) \frac{\partial u_j}{\partial q_k} = 0 \quad \text{(A-7)}$$

for all q_k in the set (q_1, ..., q_{M-1}, q_{M+1}, ..., q_{N-1}). Further, eq A-5 indicates that $\partial u_j/\partial q_k = F_{jks}$, where s corresponds to M or N, depending on whether $k < M$ or $M < k < N$, respectively. Thus, eq A-7 can be rewritten as follows:

$$\sum_{j=1}^{N_g} [U_j - (Q_1 K_{jM} + Q_2 K_{jN})] F_{jks} = \sum_{i=1}^{M-1} q_i \sum_{j=1}^{N_g} F_{jiM} F_{jks} - \sum_{i=M+1}^{N-1} q_i \sum_{j=1}^{N_g} F_{jiN} F_{jks} \quad \text{(A-8)}$$

Considering all possible q_k, eq A-8 is better represented in matrix notation as

$$\mathbf{c} = \mathbf{a} \begin{bmatrix} \mathbf{B_1} & 0 \\ 0 & \mathbf{B_2} \end{bmatrix} \quad \text{(A-9)}$$

where, $c^k = \sum_{j=1}^{N_g} [U_j - (Q_1 K_{jM} + Q_2 K_{jN}) F_{jks}]$, $\mathbf{a} = (q_1, ..., q_{M-1}, q_{M+1}, ..., q_{N-1})$, $B_1^{lk} = \sum_{j=1}^{N_g} F_{jlM} F_{jks}$, and $B_2^{lk} = \sum_{j=1}^{N_g} F_{jlN} F_{jks}$. Note that vector **c** and matrix **B** are only functions of the electrostatic potential U_j evaluated at the grid points ($j = 1 \ldots N_g$), the distances between atomic positions the grid points, r_{jk} and the partial charges Q_1 and Q_2. Therefore, the atomic charges (**a**) can be obtained by inversion of eqs A-9 and A-2.

Note Added after ASAP Publication. This article was inadvertently released ASAP on November 18, 2005 before several text corrections were made. The correct version was posted on December 5, 2005.

References

(1) van-der Vaart, A.; Bursulaya, B.; Brooks, C.; Merz, K. *J. Phys. Chem. B* **2000**, *104*, 9554−9563.

(2) Halgren, T.; Damm, W. *Curr. Opin. Struct. Biol.* **2001**, *11*, 236−242.

(3) Roux, B.; Berneche, S. *Biophys. J.* **2002**, *82*, 1681−1684.

(4) Rick, S. W.; Stuart, S. J. *Rev. Comput. Chem.* **2002**, *18*, 89−146.

(5) Ponder, J.; Case, D. A. *Adv. Prot. Chem.* **2003**, *66*, 27−47.

(6) Cieplak, P.; Caldwell, J.; Kollman, P. *J. Comput. Chem.* **2001**, *22*, 1048−1057.

(7) Banks, J. L.; Kaminski, G. A.; Zhou, R.; Mainz, D. T.; Berne, B. J.; Friesner, R. A. *J. Chem. Phys.* **1999**, *110*, 741−754.

(8) Stern, H. A.; Kaminski, G. A.; Banks, J.; Zhou, R.; Berne, B. J.; Friesner, R. A. *J. Phys. Chem. B* **1999**, *103*, 4730−4737.

(9) Gresh, N. *J. Phys. Chem. A* **1997**, *101*, 8680−8694.

(10) Masella, M.; Gresh, N.; Flament, J. P. *J. Chem. Soc., Faraday Trans.* **1998**, *94*, 2745−2753.

(11) Gresh, N.; Tiraboschi, G.; Salahub, D. R. *Biopolymers* **1998**, *45*, 405−425.

(12) Hermida-Ramon, J. M.; Brdarski, S.; Karlstrom, G.; Berg, U. *J. Comput. Chem.* **2003**, *24*, 161−176.

(13) Palmo, K.; Mannfors, B.; Mirkin, N. G.; Krimm, S. *Biopolymers* **2003**, *68*, 383−394.

(14) Palmo, K.; Krimm, S. *J. Comput. Chem.* **1998**, *19*, 754−768.

(15) Mannfors, B. E.; Mirkin, N. G.; Palmo, K.; Krimm, S. *J. Phys. Chem. A* **2003**, *107*, 1825−1832.

(16) Barnes, P.; Finney, J. L.; Nicholas, J. D.; Quinn, J. E. *Nature* **1979**, *282*, 459−464.

(17) Stillinger, F. H.; David, C. W. *J. Chem. Phys.* **1978**, *69*, 1473−1484.

(18) Corongiu, G. *Int. J. Quantum Chem.* **1992**, *42*, 1209−1235.

(19) Sprik, M.; Klein, M. L. *J. Chem. Phys.* **1988**, *89*, 7556−7560.

(20) Bernardo, D. N.; Ding, Y.; Krogh-Jespersen, K.; Levy, R. M. *J. Phys. Chem.* **1994**, *98*, 4180−4187.

(21) Yu, H.; Hansson, T.; van Gunsteren, W. F. *J. Chem. Phys.* **2003**, *118*, 221−234.

(22) Saint-Martin, H.; Hernandez-Cobos, J.; Bernal-Uruchurtu, M. I.; Ortega-Blake, I.; Berendsen, H. J. C. *J. Chem. Phys.* **2000**, *113*, 10899−10912.

(23) Stern, H. A.; Rittner, F.; Berne, B. J.; Friesner, R. A. *J. Chem. Phys.* **2001**, *115*, 2237−2251.

(24) Ren, P.; Ponder, J. W. *J. Phys. Chem. B* **2003**, *107*, 5933−5947.

(25) Burnham, C. J.; Xantheas, S. S. *J. Chem. Phys.* **2002**, *116*, 1479−1492.

(26) Perutz, M. *Science* **1978**, *201*, 1187−1191.

(27) Warshel, A. *Acc. Chem. Res.* **1981**, *14*, 284−290.

(28) Sharp, K.; Honig, B. *Annu. Rev. Biophys. Biophys. Chem.* **1990**, *19*, 301−332.

(29) Nakamura, H. *Q. Rev. Biophys.* **1996**, *29*, 1−90.

(30) Warshel, A. In *Computer modeling of chemical reactions in enzymes and solutions*; John Wiley & Sons: New York, 1991; pp 209−211, 225−228.

(31) Gascon, J.; Batista, V. *Biophys. J.* **2004**, *87*, 2931−2941.

(32) Gunner, M.; Nichols, A.; Honig, B. *J. Chem. Phys.* **1996**, *100*, 4277−4291.

(33) Parson, W. *Photosynth. Res.* **2003**, *76*, 81−92.

(34) Sham, Y.; Muegge, I.; Warshel, A. *Proteins* **1999**, *36*, 484−500.

(35) Okamura, M.; Feher, G. *Annu. Rev. Biochem.* **1992**, *61*, 861−896.

(36) Popovic, D.; Stuchebrukhov, A. *J. Am. Chem. Soc.* **2004**, *126*, 1858−1871.

(37) Burykin, A.; Warshel, A. *FEBS Lett.* **2004**, *570*, 41−46.

(38) Aqvist, J.; Luzhkov, V. *Nature* **2000**, *404*, 881−884.

(39) Bliznyuk, A.; Rendell, A.; Allen, T.; Chung, S. *J. Phys. Chem. B* **2001**, *105*, 12674−12679.

(40) Simonson, T.; Archontis, G.; Karplus, M. *J. Phys. Chem. B* **1997**, *101*, 8349−8362.

(41) Simonson, T.; Archontis, G.; Karplus, M. *J. Phys. Chem. B* **1999**, *103*, 6142−6156.

(42) Muegge, I.; Rarey, M. Small molecule docking and scoring. In *Reviews in Computational Chemistry*; Lipkowitz, K., Boyd, D., Eds.; Wiley-VCH: New York, 2001; pp 1−60.

(43) Vasilyev, V.; Bliznyuk, A. *Theor. Chem. Acc.* **2004**, *112*, 313−317.

(44) Grater, F.; Schwarzl, S.; Dejaegere, A.; Fischer, S.; Smith, J. *J. Phys. Chem. B* **2005**, *109*, 10474−10483.

(45) Cho, A.; Guallar, V.; Berne, B.; Friesner, R. *J. Comput. Chem.* **2005**, *26*, 915−931.

(46) Andre, I.; Kesvatera, T.; Jonsson, B.; Akerfeldt, K.; Linse, S. *Biophys. J.* **2004**, *87*, 1929−1938.

(47) Sheinerman, F. B.; Norel, R.; Honig, B. *Curr. Opin. Struct. Biol.* **2000**, *10*, 153−159.

(48) Ehrlich, L. Protein−Protein Docking. In *Reviews in Computational Chemistry*; Lipkowitz, K., Boyd, D., Eds.; Wiley-VCH: New York, 2001; pp 61−97.

(49) Sheinerman, F. B.; Honig, B. *J. Mol. Biol.* **2002**, *318*, 161−177.

(50) Veselovsky, A.; Ivanov, Y.; Ivanov, A.; Archakov, A.; Lewi, P.; Janssen, P. *J. Mol. Recognit.* **2002**, *15*, 405−422.

(51) Klahn, M.; Schlitter, J.; Gerwert, K. *Biophys. J.* **2005**, *88*, 3829−3844.

(52) Warshel, M. S. A. S. A. *Proc. Natl. Acad. Sci.* **2003**, *100*, 14834−14839.

(53) Nakano, T.; Kaminuma, T.; Sato, T.; Akiyama, Y.; Uebayasi, M.; Kitaura, K. *Chem. Phys. Lett.* **2000**, *318*, 614−618.

(54) Nakano, T.; Kaminuma, T.; Sato, T.; Fukuzawa, K.; Akiyama, Y.; Uebayasi, M.; Kitaura, K. *Chem. Phys. Lett.* **2002**, *351*, 475−480.

(55) Exner, T.; Mezey, P. *J. Phys. Chem. A* **2002**, *106*, 11891−11800.

(56) Gao, A.; Zhang, D.; Zhang, J.; Zhang, Y. *Chem. Phys. Lett.* **2004**, *394*, 293−297.

(57) Exner, T.; Mezey, P. *J. Phys. Chem. A* **2004**, *108*, 4301−4309.

(58) Mei, Y.; Zhang, D.; Zhang, J. *J. Phys. Chem. A* **2005**, *109*, 2−5.

(59) Li, S.; Fang, T. *J. Am. Chem. Soc.* **2005**, *127*, 7215−7226.

(60) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227−249.

(61) Halgren, T.; Damm, W. *Curr. Opin. Struct. Biol.* **2001**, *11*, 236−242.

(62) Dykstra, C. *Chem. Rev.* **1993**, *93*, 2339−2353.

(63) Wang, W.; Donini, O.; Reges, C.; Kollman, P. *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 211−243.

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

(65) Jorgensen, W.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657−1666.

(66) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.

(67) Field, M. *Mol. Phys.* **1997**, *91*, 835−846.

(68) Piquemal, J.; Williams-Hubbard, B.; Fey, N.; Deeth, R.; Gresh, N.; Giessner-Prettre, C. *J. Comput. Chem.* **2003**, *24*, 1963−1970.

(69) Piquemal, J.; Gresh, N.; Giessner-Prettre, C. *J. Phys. Chem. A* **2003**, *107*, 10353−10359.

(70) Maple, J.; Cao, Y.; Damm, W.; Halgren, T.; Kaminski, G.; Zhang, L.; Friesner, R. *J. Chem. Theory Comput.* **2005**, *1*, 694−715.

(71) York, D.; Lee, T.; Yang, W. *J. Am. Chem. Soc.* **2005**, *127*, 7215−7226.

(72) Stewart, J. *Int. J. Quantum Chem.* **1996**, *58*, 133−146.

(73) Stewart, J. Mopac2000; Fujitsu Ltd.: Tokyo, 1999.

(74) Cummins, P.; Gready, J. In *Hybrid Quantum Mechanical and Molecular Mechanical Methods*; Gao, J., Thompson, M., Eds.; American Chemical Society: Washington, DC, 1998; pp 250−263.

(75) Zuegg, J.; Bliznyuk, A.; Gready, J. *Mol. Phys.* **2003**, *101*, 2437−2450.

(76) Titmuss, S.; Cummins, P.; Rendell, A.; Bliznyuk, A.; Gready, J. *J. Comput. Chem.* **2002**, *23*, 1314−1322.

(77) Titmuss, S.; Cummins, P.; Bliznyuk, A.; Rendell, A.; Gready, J. *Chem. Phys. Lett.* **2000**, *320*, 169−176.

(78) Williams, D. Net atomic charge and multipole models for ab initio molecular electric potential. In *Reviews in Computational Chemistry*; Lipkowitz, K., Boyd, D., Eds.; Wiley-VCH: New York, 1991; pp 219−271.

(79) Stewart, J. Semiempirical molecular orbital methods. In *Reviews in Computational Chemistry*; Lipkowitz, K., Boyd, D., Eds.; Wiley-VCH: New York, 1990; pp 45−81.

(80) Zerner, M. Semiempirical molecular orbital methods. In *Reviews in Computational Chemistry*; Lipkowitz, K., Boyd, D., Eds.; Wiley-VCH: New York, 1991; pp 313−365.

(81) Repasky, M.; Chandrasekar, J.; Jorgensen, W. *J. Comput. Chem.* **2002**, *23*, 1601−1622.

(82) Maseras, M.; Morokuma, K. *J. Comput. Chem.* **1995**, *16*, 1170−1179.

(83) Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. *J. Phys. Chem.* **1996**, *100*, 19357−19363.

(84) Humbel, S.; Sieber, S.; Morokuma, K. *J. Chem. Phys.* **1996**, *16*, 1959−1967.

(85) Dapprich, S.; Komaromi, K.; Byun, K.; Morokuma, K.; Frisch, M. *J. Mol. Struct. (Theochem)* **1999**, *461*, 1−21.

(86) Vreven, T.; Morokuma, K. *J. Comput. Chem.* **2000**, *21*, 1419−1432.

(87) Vreven, T.; Mennucci, B.; daSilva, C. O.; Morokuma, K.; Tomasi, J. *J. Chem. Phys.* **2001**, *115*, 62−72.

(88) Vreven, T.; Morokuma, K. *Theor. Chem. Acc.* **2003**, *109*, 125−132.

(89) 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 A.1*; 2003.

(90) The choice of basis set and level of theory aims to make contact with earlier work, where the HF/6-31G* level of theory has been the standard for calculations of ESP atomic charges. For example, Peter Kollman and co-workers obtained ESP and RESP charges in the early parametrization of Amber (Cornell et al. *J. Am. Chem. Soc.* **1993**, *115*, 9620; *J. Am. Chem. Soc.* **1995**, *117*, 5179), and they showed that the HF/6−3lG* level was able to reproduce experimental free energies of solvation. In the particular case of ion channels we used this level of theory to present a consistent comparison to the electrostatic potential provided by Amber. However, it is important to note that the proposed methodology is not limited to any particular basis set or level of quantum chemistry.

(91) Politzer, P.; Murray, J. Molecular Electrostatic Potentials and Chemical Reactivity. In *Reviews in Computational Chemistry*; Lipkowitz, K., Boyd, D., Eds.; Wiley-VCH: New York, 1991; pp 273−312.

(92) Murray, J.; Politzer, P. Electrostatic Potentials: Chemical Applications. In *The Encyclopedia of Computational Chemistry*; Schleyer, P., Allinger, N., Clark, T., Gasteiger, J., Kollman, P., Schaefer, H., Schreiner, P., Eds.; Wiley & Sons: Chichester, U.K., 1998; pp 912−920.

(93) Murray, J.; Politzer, P. The Molecular Electrostatic Potential: A Tool for Understanding and Predicting Molecular Interactions. In *Molecular Orbital Calculations for Biological Systems*; Sapse, A.-M., Ed.; Oxford University Press: New York, 1998; pp 49−84.

(94) Oprea, T.; Waller, C. Theoretical and Practical Aspects of Three-Dimensional Quantitative Structure-Activity Relationships. In *Reviews in Computational Chemistry*; Lipkowitz, K., Boyd, D., Eds.; Wiley-VCH: New York, 1997; pp 127−182.

(95) Greco, G.; Novellino, E.; Martin, Y. Approaches to Three-Dimensional Quantitative Structure-Activity Relationships. In *Reviews in Computational Chemistry*; Lipkowitz, K., Boyd, D., Eds.; Wiley-VCH: New York, 1997; pp 183−240.

(96) Vdaykumar, S.; Bugg, C.; Cook, W. *J. Mol. Biol.* **1987**, *194*, 531−544.

(97) Betzel, C.; Pal, G.; Saenger, W. *Acta Crystallogr. B* **1988**, *44*, 163−172.

(98) Sasaki, K.; Dockerill, S.; Adamiak, D.; Tickle, I.; Blundell, T. *Nature* **1975**, *257*, 751−757.

(99) Doyle, D. A.; Cabral, J. M.; Pfuetzner, R. A.; Kuo, A.; Gulbis, J. M.; Cohen, S. L.; Chait, B. T.; MacKinnon, R. *Science* **1998**, *280*, 69−77.

(100) Elcock, A.; Sept, D.; McCammon, J. *J. Phys. Chem. B* **2001**, *105*, 1504−1518.

(101) Honig, B.; Nicholls, A. *Science* **1995**, *268*, 1144−1149.

(102) Honig, B.; Hubbell, W. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 5412−5416.

(103) Hendsch, Z.; Tidor, B. *Protein Sci.* **1994**, *3*, 211−226.

(104) Yang, A.-S.; Honig, B. *J. Mol. Biol.* **1995**, *252*, 366−376.

(105) Yang, A.-S.; Honig, B. *J. Mol. Biol.* **1995**, *252*, 351−365.

(106) Schapira, M.; Totrov, M.; Abagyan, R. *J. Mol. Recognit.* **1999**, *12*, 177−190.

(107) Luo, R.; David, L.; Hung, H.; Devaney, J.; Gilson, M. *J. Phys. Chem. B* **1999**, *103*, 727−736.

(108) Froloff, N.; Windermuth, A.; Honig, B. *Protein Sci.* **1997**, *6*, 1293−1301.

(109) Chong, L.; Dempster, S.; Hendsch, Z.; Lee, L.; Tidor, B. *Protein Sci.* **1998**, *7*, 206−210.

(110) Lee, L.-P.; Tidor, B. *Protein Sci.* **2001**, *10*, 362−377.

(111) Lee, L.-P.; Tidor, B. *Nat. Struct. Biol.* **2001**, *8*, 73−76.

(112) Dong, F.; Vijayakumar, M.; Zhou, H. *Biophys. J.* **2003**, *85*, 49−60.

(113) Wang, T.; Tomic, S.; Gabdoulline, R.; Wade, R. *Biophys. J.* **2004**, *87*, 1618−1630.

(114) Warwicker, J.; Watson, H. *J. Mol. Biol.* **1982**, *157*, 671−679.

(115) Gilson, M.; Sharp, K.; Honig, B. *J. Comput. Chem.* **1987**, *9*, 327−335.

(116) Nicholls, A.; Honig, B. *J. Comput. Chem.* **1991**, *12*, 435−445.

(117) Nina, M.; Beglov, D.; Roux, B. *J. Phys. Chem.* **1997**, *101*, 5239−5248.

(118) Schreiber, G.; Fersht, A. R. *Biochemistry* **1993**, *32*, 5145−5150.

(119) Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A.; Cao, Y. X.; Murphy, R. B.; Zhou, R.; Halgren, T. A. *J. Comput. Chem.* **2002**, *23*, 1515−1531.

CT050218H

JCTC Journal of Chemical Theory and Computation

Base Flipping in a GCGC Containing DNA Dodecamer: A Comparative Study of the Performance of the Nucleic Acid Force Fields, CHARMM, AMBER, and BMS

U. Deva Priyakumar and Alexander D. MacKerell, Jr.*

Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, Baltimore, Maryland 21201

Received August 5, 2005

Abstract: The improving quality of empirical force field parameters along with other methodological improvements and ever increasing computational resources have lead to more reliable computations on biological macromolecules. In the case of oligonucleotides, three force fields, namely CHARMM27, AMBER4.1, and BMS, have been developed and are widely used by the simulation community. Testing of these force fields to date has primarily focused on their treatment of the canonical forms of DNA and RNA. However, many biological functions of oligonucleotides involve significant variation of their structures from the canonical forms. In the present work, the three force fields are evaluated via computation of potentials of mean force (PMF) of the base flipping process in a DNA dodecamer, 5′-GTCAGCGCATGG-3′. Results are compared with available experimental data on the equilibrium between the opened and closed (i.e. Watson−Crick base paired) state of the underlined C and its WC partner G. Quantitative analysis shows CHARMM to be in the best agreement with experiment, closely followed by AMBER with BMS in the poorest agreement. Various components contributing to the change in the free energy such as base pair interactions, stacking interactions, solvation effects, and intrinsic potential energy changes were evaluated and compared. The results indicate that while all three force fields reasonably represent the canonical structures, the balance of forces contributing to their structural and dynamic properties differ significantly.

Introduction

Molecular dynamics simulations play a dominant role in understanding the relationships among structure, energetics, and function of biological macromolecules.[1−3] While these calculations are helpful in explaining various experimental observations, they are indispensable in investigating properties that are otherwise difficult or inaccessible to experiments, including high-energy states sampled during conformational transitions. These attributes include the ability to obtain energetic information on conformational transition and relate that information to structural properties at an atomic level of detail. Accordingly, an important consideration when applying MD simulations to biological macromolecules is the quality of the empirical force field being used to correctly represent the relationship between structure and energetics. The past decade has witnessed tremendous progress in the development of these force fields, thereby enabling more reliable computations on biomolecules in general and nucleic acids in particular.[3−12] Various force fields optimized for nucleic acids are available, including the CHARMM27,[13,14] AMBER4.1,[15,16] and Bristol-Myers-Squibb (BMS)[17] all-atom force fields. To date, tests of these force fields have focused on the canonical structures of DNA and RNA. These tests have indicated that the above force fields satisfactorily treat the canonical structures, although limitations in each of the force fields have been noted.[17−21] In the present paper we extend the tests of these force fields to include a conforma-

* Corresponding author phone: (410)706-7442; fax: (410)706-5017; e-mail: amackere@rx.umaryland.edu. Corresponding author address: 20 Penn Street, Baltimore, MD 21201.

10.1021/ct0501957 CCC: $33.50
Published on Web 11/16/2005

tional transition to a noncanonical conformation of DNA, namely base flipping.

Base flipping is a process by which one of the bases of the DNA is displaced from its base paired state, moving out of the double helix typically leaving its WC-base paring counterpart in its original position.[22−26] Such a process, though energetically unfavorable, is favored during interactions with selected proteins, which assist flipping of the base to perform chemical reactions on the otherwise inaccessible base.[23,25,27−29] Base flipping is also adopted by transcription proteins in order to achieve stable protein−DNA complexes.[22,30] In the absence of a protein, nucleic acids undergo base opening whose dynamics have been extensively studied using NMR imino proton exchange experiments on various sequences.[31−35] These experiments yield base opening rates along with the equilibrium between the open and closed states assuming a two-state model. Taking advantage of these data, a direct comparison of the equilibrium between the open and closed states from potential of mean force (PMF) calculations based on MD simulations has been performed.[36]

The present study focuses on evaluation of the CHARMM-27, AMBER (Parm94), and BMS nucleic acid force fields in modeling base flipping in a DNA dodecamer, 5′-GTCAGCGCATGG-3′. PMF calculations for the flipping of the underlined C and its WC base paired counterpart, G, were performed. Equilibrium constant for the base opening/closing process were calculated using the PMFs generated with all the force fields, and the results were compared with the available experimental data. The results are presented and discussed in the following order: Justification of the current methodology and the adequacy of the length of the simulation are discussed followed by presentation of the free energy profiles corresponding to base flipping computed using the three force fields. This is followed by the comparison of the theoretical and experimental data of the equilibrium between the closed and open states. Finally, various factors that affect the base flipping process such as disruption of base pairing and stacking interactions, solvation effects, and intrinsic energetic properties of the DNA are discussed.

Methods

All calculations were performed using the CHARMM program.[37,38] Three different force fields were employed, CHARMM27,[13,14] AMBER4.1 (PARM94),[16] and Bristol-Myers Squibb (BMS).[17] Initially, coordinates for the DNA in the canonical B-form were generated using QUANTA[39] and overlaid onto a preequilibriated solvent box containing sodium ions. The solvent shell extended approximately 8 Å beyond the DNA along the helical axis and 20 Å perpendicular to the axis. Those solvent molecules or the sodium ions whose non-hydrogen atom were within 1.8 Å of non-hydrogen atoms of the DNA were removed, and then the number of the sodium atoms was adjusted to attain electrical neutrality. The systems were minimized for 500 Adopted-Basis Newton Rapheson (ABNR) steps with harmonic constrains of 2.0 kcal/mol/Å on the non-hydrogen atoms of the DNA followed by a 20 ps molecular dynamics simulation in the NVT ensemble. The CRYSTAL[40] module in CHARMM was used for the periodic boundary conditions, and electrostatic interactions were treated using the Particle Mesh Ewald method.[41] Real space electrostatic and Lennard-Jones cutoffs were 12 Å with a switch smoothing function from 10 to 12 Å for the LJ term. The nonbond list was maintained to 14 Å and heuristically updated. The final configurations from the NVT simulation were subjected to minimization for 500 ABNR steps and then a 500 ps NPT MD simulation without any constraints. The final conformations from the 500 ps simulation were taken for the PMF calculations.

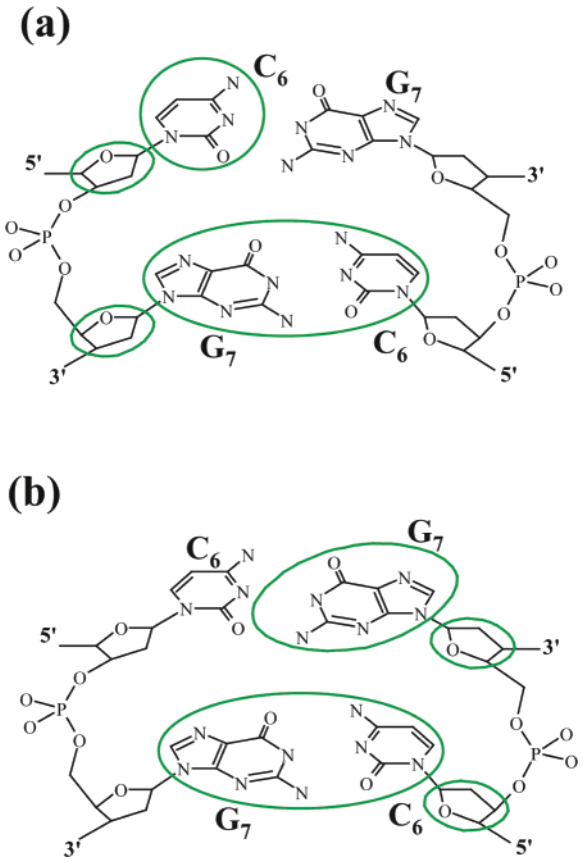


Figure 1. Schematic representation of the center of mass pseudodihedral angle for target C- and G-base flipping (a and b).[36] The dihedral angle formed by the centers of mass of the atoms in the four circles is termed the COM pseudodihedral angle. Values of ∼10° and ∼30° correspond to the WC base paired state in the free energy profiles for C- and G-base flipping, respectively.

Base flipping PMFs were obtained by performing 72 independent MD simulations (i.e. windows) with different pseudodihedral center of mass (COM) restraints (Figure 1) in 5° increments from 0 to 360° from which the probability distributions were obtained for calculation of the free energy surfaces.[36] The COM pseudodihedral angle is defined by the dihedral angle formed by four coordinates defined based on the centers of mass of four sets of atoms: (a) the GC base pair 3′ to the flipping base, (b) the sugar attached to the adjacent base 3′ to the flipping base, (c) the sugar attached to the flipping base, and (d) the flipping base (Figure 1). Initial conformations corresponding to the 72 flipped states for the target C and G bases being flipped were generated as previously described.[36] Briefly, an initial structure corresponding to the first window ($x = 0°$) was obtained using a

0.5 ps simulation with a harmonic potential (force constant = 10 000 kcal/mol/rad^2) on the COM pseudodihedral angle. For the remaining windows, the flipped conformers along both grooves were obtained by performing a series of 0.5 ps MD simulations in the presence of the harmonic potential incremented by ±5° from the final structure from the previous window. This was repeated via both grooves out to 180° yielding the 72 starting structures. The resulting coordinates corresponding to the 72 different flipped conformations of the DNA were then overlaid onto a water sphere of radius 35 Å. The solvent molecules whose non-hydrogen atom was within 1.8 Å of any non-hydrogen atom of the oligonucleotide were deleted, and the number of sodium ions was adjusted to attain electrical neutrality. The resulting systems were then subjected to a 500 step steepest descent minimization, and equilibration of each window was done for 60 ps followed by a 160 ps production run. Nonbond interactions were treated via atom based truncation with the nonbonded lists updated heuristically with a list cutoff of 14 Å, a nonbond cutoff of 12 Å, and the smoothing functions initiated at 10 Å. Electrostatic and LJ interactions were smoothed using the force shift and force switch methods, respectively.[42] An integration time step of 2 fs, a temperature of 300 K, and SHAKE[43] to constrain the covalent bonds involving hydrogen atoms were used during the NVT simulation applying the Nosé-Hoover temperature coupling scheme.[44] During the minimization, equilibrium, and production runs, the following restraints were imposed: (a) the terminal base pairs of the DNA were harmonically restrained to their initial spatial coordinates using a force constant of 2.0 kcal/mol/Å; (b) water density of the systems was maintained by using the mean field solvent boundary potential included in the miscellaneous mean field potential (MMFP) module in CHARMM;[45] and (c) a harmonic umbrella potential, $w_i(x) = k_i (x - x_i)^2$ (k_i is the force constant, 1000 kcal mol^{-1} rad^{-2}; x is the center of mass (COM) dihedral angle; and x_i is the restrained value of the angle) was used for the COM pseudodihedral angle. The value for the pseudodihedral angle was recorded every time step during the simulation for obtaining the probability distributions; other analyses were performed on time frames recorded every 1 ps of the trajectories. PMFs were obtained using the weighted histogram analysis method (WHAM) procedure that enforces periodicity of the reaction coordinate,[46,47] with a width of 0.5° for the pseudodihedral angle as previously described. Stacking interaction of the flipping base with its neighbors were calculated by considering both electrostatic and van der Waals terms using the INTER command implemented in CHARMM.[48] The neighboring bases immediate to the flipping base and their base pair counterparts in the complementary strand were considered for these calculations using the real space nonbond interaction cutoffs listed above. Calculation of the stacking interactions involved only the specified nucleic acid bases and not the sugar or phosphate groups.

Results and Discussion

Adequacy of the Sampling. Essential for the validity of the theoretical-experimental quantitative comparison is the convergence of the obtained free energy profiles; hence, the adequacy of the length of the 72 simulations comprising the PMF was first verified. For each free energy profile generated, each simulation (i.e. window) was performed for 220 ps, yielding a sum for all 72 simulations defining each PMF of 4.3 ns for the equilibration and 11.5 ns for the production runs. MacKerell and co-workers have performed limited tests on the convergence issues with respect to the length of the simulation and found that 220 ps MD simulation (i.e. 60 ps equilibration plus 160 ps production) for each window is long enough for satisfactory convergence.[49] To further validate the adequacy of the length of the simulations for the three force fields, the PMFs with respect to the length of the simulation were calculated for every 20 ps range from 60 to 220 ps (Figure S1 in the Supporting Information). The free energy profile obtained for the whole 60−220 ps range is also given for comparison. The PMFs overlap after the initial 60−80 ps sampling periods fluctuating around the 60−220 ps surfaces. In addition, the 72 MD simulations that used the CHARMM force field were each extended to 440 ps. The free energy profiles calculated from the 60−220 ps, 220−440 ps, and 60−440 ps windows are given in Figure 2. Comparison of the PMFs from these sampling ranges shows only minor differences with respect to the increase in sampling. While not absolute proof, behavior of the PMFs strongly suggests that they are adequately converged at 220 ps to allow for quantitative analysis and detailed structural and energetic analysis of the flipping profiles.

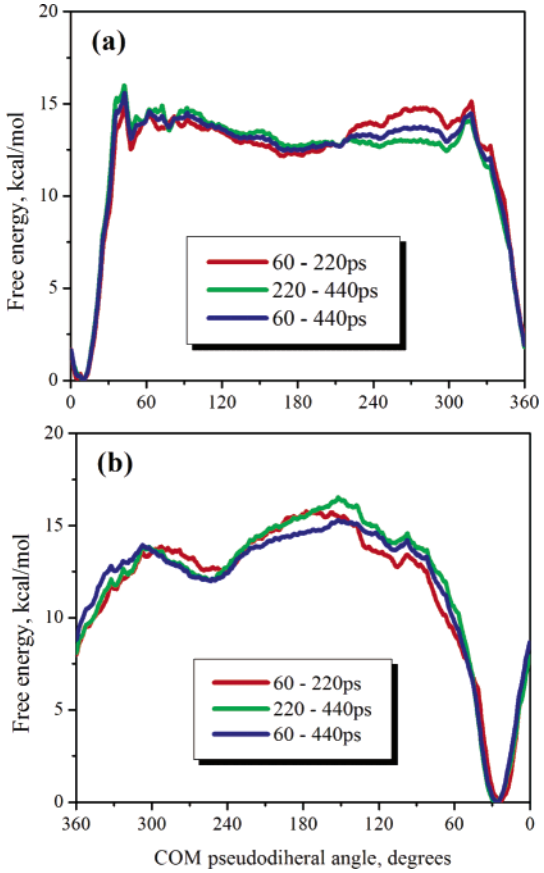


Figure 2. PMFs obtained for the C- (a) and G-base (b) flipping using the 60−220, 220−440, and 60−440 ps windows of the MD simulations using the CHARMM force field.

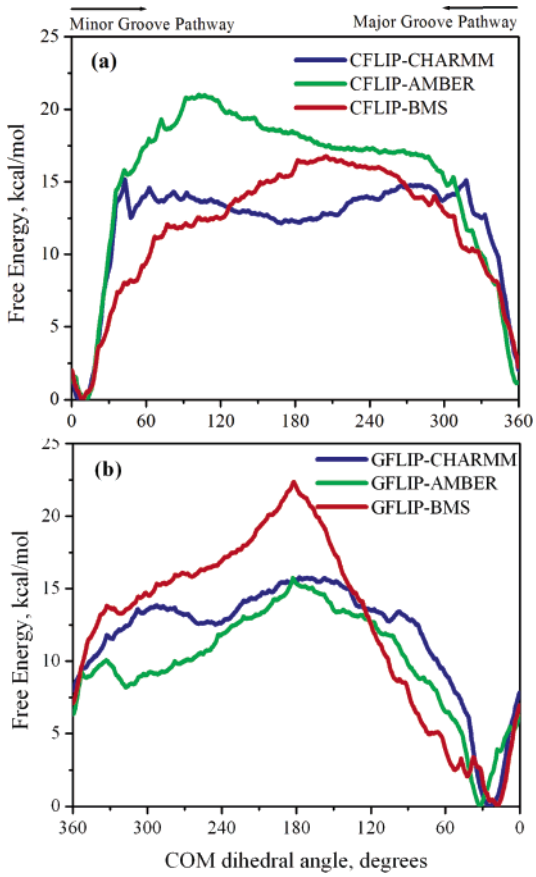


Figure 3. Free energy profiles obtained using CHARMM (blue), AMBER (green), and BMS (red) force fields as a function of the COM pseudodihedral angle (Figure 1) for C-base flipping (a) and G-base flipping (b). The free energies were calculated using the 60−220 ps windows of the MD simulations.

Free Energy Profiles. The PMFs calculated for the target C- and G-base flipping are depicted in Figure 3 (parts (a) and (b), respectively) for the CHARMM, AMBER, and BMS force fields. Inspection of these figures indicates that the free energy profiles predicted using the three force fields are quite different both qualitatively and quantitatively. In the case of C-flipping, the WC base paired state at a pseudodihedral angle of ∼10° is predicted to be the minimum energy structure by all the three force fields (11, 11, and 9° by CHARMM, AMBER, and BMS, respectively). However, the shapes of the profiles and the barrier heights for base flipping give rise to three different scenarios. The free energy profiles obtained using CHARMM and AMBER indicate that the energy increases sharply when the target C base flips out of DNA duplex from both the major and minor grooves, whereas with BMS, the increase in energy is gradual especially along the minor groove pathway. The CHARMM PMF exhibits a broad, shallow minimum of approximately 2.5 kcal/mol in the range of 60−320° for the C-base flipped state with an energy of approximately 12 kcal/mol above the global minimum. The energy barrier for flipping from the minor and major grooves is around 16 kcal/mol. This is consistent with the previous computational and experimental studies, which predict that the base flipping via both the minor and major grooves is feasible.[36,50−54] In contrast to CHARMM, the other two force fields do not have a minimum for the flipped out state. The flipping profile obtained with AMBER indicates that the minor groove pathway is energetically more expensive by about 4 kcal/mol than the major groove pathway. This is consistent with the conventional wisdom that for the base to move out via the minor groove, it has to overcome steric effects from the close lying backbone compared to the major groove side. BMS, on the other hand, predicts a more gradual increase of energy via the minor groove and a sharp increase via the major groove, which indicates that the former pathway is slightly preferred over the other. With both AMBER and BMS single distinct maxima occur in the PMF at ∼100 and ∼200°, respectively. Irrespective of the force field employed, the results indicate that the change in free energy among the various flipped states (COM pseudodihedral angle range of ∼60 to 300°) is not drastic. This implies that once the base flips out, it is expected to sample a wide range of conformations.

Results for G-flipping for the 3 force fields are presented in Figure 3(b). The pseudodihedral angle at which the WC base pairing occurs for G-flipping is different from that in the C-flipping profiles based on the definitions of the flipping pseudodihedral angle as is the direction of the rotation corresponding to minor versus major groove flipping.[36] Accordingly, the *x*-axis is reversed to allow for visualization of the minor and major flipping pathways to coincide for C- and G-flipping. The free energy profiles calculated using the three force fields show COM pseudodihedral angles at which WC base paired states occur slightly deviate from each other (25° for CHARMM, 32° for AMBER, and 18° for BMS). This may be due to slight deviations in relative orientations of the sugar and the adjacent GC base pairs, which are used to define the pseudodihedral angle. Qualitatively, the change in the free energies with respect to the pseudodihedral angle computed by CHARMM and AMBER are similar, especially via the major groove pathway. Both AMBER and BMS have distinct maxima at ∼170°, with BMS lacking any significant local minima for the flipped state; such states are seen with CHARMM and AMBER at approximately 240 and 310°, respectively. However, these minima are shallow, being ∼2 kcal/mol deep. Overall, it is evident that while all three force fields show the distinct minima associated with the WC base paired states, there are significant qualitative differences between the models.

Imino proton exchange studies on a GCGC sequence indicate the presence of local minima for the fully flipped state which lies about 9 kcal/mol above the WC base paired state.[32] In the free energy profiles for C-flipping with CHARMM and G-flipping for both CHARMM and AMBER local minima are present, though these minima are shallow. Experimental studies have shown that the lifetime for the WC base paired state is in the order of milliseconds, and the proposed base open state corresponds to a metastable state with a lifetime in the nanosecond range.[33,55,56] Based on the difference of approximately 10^6 between the lifetimes of the WC and flipped minima, the difference between the barriers corresponding to base closing and opening processes is calculated to be about 14 kcal/mol according to the

transition state theory and assuming that the preexponential contributions are identical for the two processes. In the present study, CHARMM predicts a shallow minimum for the flipped out state for both C- and G-flipping. The differences in the barrier corresponding to base opening and closing calculated using the CHARMM free energy profiles are 12.1 and 11.8 kcal/mol for C- and G-flipping, respectively, which is consistent with the experimental results.

Comparison of Experimental and Calculated Equilibria between the Open and Closed States. Base flipping leads to exposure of the imino proton of the bases, which are otherwise hidden in the DNA duplex, to the solvent environment. Upon exposure, the imino protons from G-H1 or T(U)-H3 undergo exchange with the solvent. This process has been extensively used to measure the opening and closing rates of the bases and the equilibrium between the open and closed states in nucleic acids.[31−33,57,58] Quantitative analysis of these experiments is based on a two-state model where the equilibrium between the two states is studied within the assumption that the imino protons in the closed state are not accessible for exchange. Experimentally, it has been observed that the base pairing and opening process occurs on the millisecond time scale with the equilibrium between the open and closed states typically in the range of 10^{-7}. In particular, experiments on DNA containing a central GCGC sequence have yielded an equilibrium constant of 3.3×10^{-7}.[32] This value may be used for quantitative evaluation of the present PMFs, as previously performed.[36] It should be noted that estimates of the free energy of opening have been made based on the measured opening rates. However, exact calculation of the activation free energy of opening requires knowledge of the preexponential term in transition state theory.[59,60]

To calculate equilibrium constants, the 72 windows have to be assigned to open or closed states, following which summation over the probabilities for the two states allows for calculation of the equilibrium constants. Base open states are defined as those conformations whose imino proton is accessible for exchange with solvent, though they may partly be base paired. To identify the windows that comprise the open state, the solvent accessible surface area[61] of the N1 and H1 atoms of the guanine base was calculated using a probe radius of 1.4 Å with an accuracy of 0.01 Å, with those windows having an accessibility greater than zero assigned as being open. To obtain the probabilities of each state, the PMFs were converted to probability distributions based on a Boltzmann distribution. The mean solvent accessibilities and the probabilities as a function of the COM pseudodihedral angles for the C- and G-flipping obtained using CHARMM, AMBER, and BMS are depicted in Figure 4. Expectedly, the solvent accessibility of the base open and closed states differ significantly in the case of C- versus G-flipping due to differential exposure of the G(H1) imino proton to the environment. For C-flipping (Figure 4a−c), the G base stays in the duplex and hence the solvent accessibility difference is not large; however, the imino proton is accessible for exchange when the C-base flips out as reflected in the variation of the solvent accessibilities as a function of COM dihedral angle.Based on the solvent accessibilities, the conformers were grouped into open and closed states based on the COM dihedral angle as listed in Table 1. The equilibrium constants calculated by integrating the unbiased probability distributions over the open and closed states along with the experimental data are also given in Table 1. The equilibrium constants for the C- and G-base opening have to be summed as the experimental data corresponds to both C- or G-base opening. From the results it is evident that CHARMM yields the best agreement with experiment, followed by AMBER with BMS in relatively poor agreement. These observations hold when variations in the windows selected for calculation of the PMF are tested, as shown in Tables S1−S3 of the Supporting Information. With both AMBER and BMS the calculated equilibrium constants are larger than the experimental values. This indicates that the open states are more favored in the force fields as compared to the experimental regimen. To better understand this behavior as well as compare how the various components of the force fields contribute to the calculated equilibria and PMFs, analysis of different structural and energetic terms as a function of the COM pseudodihedral was undertaken.

Potential Energy Contributions to the Base Flipping Free Energy Profiles. To better understand the atomistic contributions to the flipping PMFs, changes in potential energies as a function of the flipping free energy surfaces were obtained. Energetic contributions analyzed included the interaction energy between the flipping base and its WC base pairing partner, stacking interactions of the flipping base with its neighbors, interaction of the flipping base with the remainder of the DNA, interaction energies with the solvent and intrinsic energetics of the DNA itself. The initial analysis involved looking at changes in the interaction energy of the flipping base with the remainder of the DNA and with the solvent environment. Presented in Table 2 are the energy differences for the two terms between the WC states and the flipped states, where the values for the flipped states are averages over windows 180−210°, inclusive. Table S4 of the Supporting Information includes the average values for the two states, and Figure S2 shows the changes as a function of the extent of flipping. As may be seen, upon flipping the interaction energy of the base with the remainder of the DNA become less favorable due to the expected decrease in the favorable interactions between the flipping base and the DNA, with that loss of energy being similar for CHARMM and AMBER, while the value with BMS is larger. Opposing the loss of base−DNA interactions are gains in the energy of solvation of the tribase; in this case the CHARMM and AMBER values are significantly more favorable than that observed with BMS. Analysis of the magnitudes of the flipping base−DNA and solvation terms shows them to be larger than the free energy differences of approximately 15 kcal/mol for both C- and G-flipping (Figure 3). Thus, the present results indicate that the free energy of flipping is associated with large interactions of the flipping base with the remaining DNA and with the solvent environment, with those contributions acting to compensate for each other, yielding a smaller free energy difference than those components themselves. In addition, it is clear that differences in these terms exist between the force fields; additional

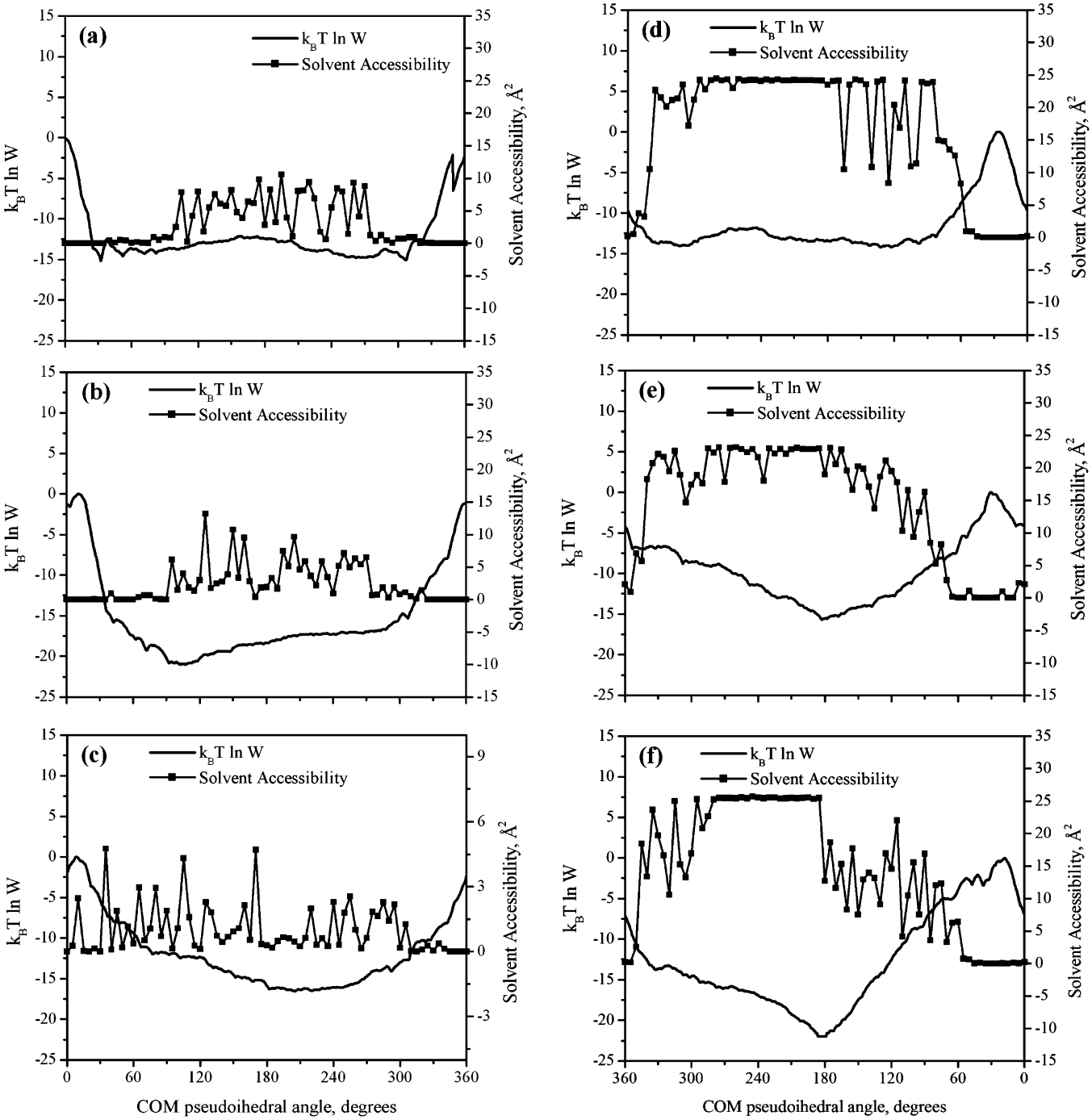


Figure 4. Solvent accessibilities (filled circles) of the imino proton of the orphan G averaged over the final 160 ps of the MD simulations and the Boltzmann weighted logarithm of the unbiased probability densities (solid line) obtained for C- (a, b, c) and G-base flipping (d, e, f) using CHARMM (a and d), AMBER (b and e), and BMS (c and f).

analysis will focus on obtaining a more detailed understanding of the contributions to these differences.

Base Pairing Interactions and Watson−Crick Hydrogen-Bonding Distances. The most obvious consequence to the DNA when one of the bases flips is the disruption of the hydrogen bonds formed in the native WC base paired state. The correlation of the N3(G)−N1(C) distance and the GC base pair interaction energy averaged over the final 160 ps with respect to the COM dihedral angle are given in Figure 5; the PMFs obtained are also depicted for comparison. Notable is the significant difference in the total GC interaction energy for the three force fields. The base pair interaction energies are offset to the value corresponding to the WC base paired state. The average GC interaction energies at the WC base paired states are computed to be 21.9, 26.3, and 26.2 kcal/mol using CHARMM, AMBER and BMS, respectively. These values are similar to GC base pair interaction energies published in the literature,[13,17,62,63] though differences exist due to subtle changes in the geometry of the DNA and methodological differences. Notably, both AMBER and BMS force fields significantly overestimate the interaction energies as judged by both quantum mechanical and experimental data.[13,63,64]

Table 1: Equilibrium Constants for the Equilibrium between the Base Open and Closed States Calculated Using CHARMM, AMBER, and BMS along with the Experimental Value[a]

method	C-flip	G-flip	total
CHARMM			
open states	75−320°	55−345°	
equilibrium constant	1.1×10^{-8}	3.6×10^{-7}	3.8×10^{-7}
AMBER			
open states	60−325°	65−355°	
equilibrium constant	4.6×10^{-9}	7.1×10^{-6}	7.1×10^{-6}
BMS			
open states	60−300°	70−350°	
equilibrium constant	6.1×10^{-8}	2.1×10^{-4}	2.1×10^{-4}
exptl equilibrium constant			3.3×10^{-7}

[a] Included are the ranges of the COM dihedral angles of the conformers considered as base open states used for calculation of the equilibrium constants. Open states defined based on the COM pseudodihedral angles. Closed states are all states not defined as open states.

Table 2: Differences in Interaction Energies between the WC and Fully Flipped States for the Flipping Base with the Remainder of the DNA and for the Flipping Base, the Central Tribase (GCG), and the Backbone of the Tribase with the Solvent and Their Solvent Accessible Surface Areas Obtained using the CHARMM, AMBER, and BMS Force Fields[a]

	CHARMM	AMBER	BMS
base to DNA interaction energy			
C-flip	37.1	35.3	51.0
G-flip	51.4	55.9	65.7
base to solvent interaction energy			
C-flip	−23.7	−24.6	−37.7
G-flip	−39.2	−41.0	−52.9
backbone to solvent interaction energy			
C-flip	−23.4	−18.3	0.1
G-flip	−25.8	0.5	−3.1
tribase to solvent interaction energy			
C-flip	−80.8	−79.2	−47.8
G-flip	−95.3	−90.2	−48.1
base solvent accessible surface area			
C-flip	17.0	19.2	5.7
G-flip	21.6	20.8	7.0
backbone solvent accessible surface area			
C-flip	−63.5	−14.9	−24.5
G-flip	−31.1	36.4	−79.1
tribase solvent accessible surface area			
C-flip	245.2	276.9	82.2
G-flip	247.4	239.5	57.9

[a] Interaction energies are given in kcal/mol and solvent accessibilities in Å^2. Individual energies along with error estimates from which the differences in this table were calculated are presented in Table S4 of the Supporting Information.

Expectedly, the N1−N3 distance increases and the GC base pairing interaction energy decreases in all the cases when the COM dihedral does not correspond to the WC base paired state. The change in the N1−N3 distance correlates well with the base pair interaction energies and also with the free energy changes. For C-flipping via the minor groove, the base pairing interaction is maintained up to approximately 40° (35° in case of BMS) beyond the WC base paired state with a drastic decrease in the interaction energy at this angle when going further from the WC base paired state. Interestingly, the barrier for C-flipping via the minor groove occurs at the position that the N1−N3 distance shows a marked increase with CHARMM and AMBER. Whereas from the major groove CHARMM predicts a gradual decrease in the interaction energy (i.e., becomes less favorable) and an increase in the N1−N3 distance for the pseudodiheral angle from 10 though 0° down to 315°. AMBER and BMS predict a similar change in these terms but with sudden jumps between; this behavior may contribute to the more gradual increase in the free energy profiles observed in that region for those force fields.

In the case of G-flipping (Figure 5d−f), the change in the base pair interaction as a function of the COM dihedral angle is similar in the sense that the decrease is sudden from the minor groove and gradual from the major groove. Interestingly, base paring is well maintained out to 65° in the case of AMBER as reflected in both the N1−N3 distance and interaction energy; this may be due to the overestimation of base pairing energies, which makes the orphan C base move with the flipping G base. In general, during G-flipping the orphan C base is pushed out of the DNA duplex and moves with the flipping G base along both grooves. In contrast, C-flipping requires the orphan G base to be pushed out only via the minor groove. This can be explained based on the size of the flipping base and the steric constraint via the minor groove pathway.

Stacking Interactions. Inter- and intrastrand stacking interactions of a given base with its neighbors contribute to the overall stability of the oligonucleotides.[48,65−69] During base flipping, the π-stacking stabilizing interaction of the flipping base with its neighbors is expected to change vastly; hence, this could be one of the major factors influencing the base flipping process. The average stacking interactions of the flipping C and G bases with their neighbors are depicted in Figure 5 as a function of the pseudodihedral angle. The stacking interactions of the orphan base with its neighbors were also calculated; however, we observed no appreciable variation with respect to the pseudodihedral angle. In the case of C-flipping, the stacking interactions do not start diminishing significantly via the minor groove as flipping initially proceeds from the WC base paired state (up to 75, 90, and 65° using CHARMM, AMBER, and BMS, respectively). Unexpectedly, the interaction is considerably more favorable in this region compared to that at the base paired state. This is due to the method used to calculate base stacking, such that the formation of hydrogen bonds between the flipping base with the neighboring bases that occur as the plane of the flipping base no longer stays parallel to those of the other bases, contribute to the stacking energy. Examples of such interactions with the CHARMM force field are shown in Figure 6. With AMBER this effect is significant, being <-10 kcal/mol as compared to that observed in the WC base paired state. Interestingly, this occurs despite the C stacking energy being the least favorable in AMBER as compared to CHARMM and BMS. With all three force fields the favorable stacking energy is maintained during minor groove flipping to larger pseudodihedral angles

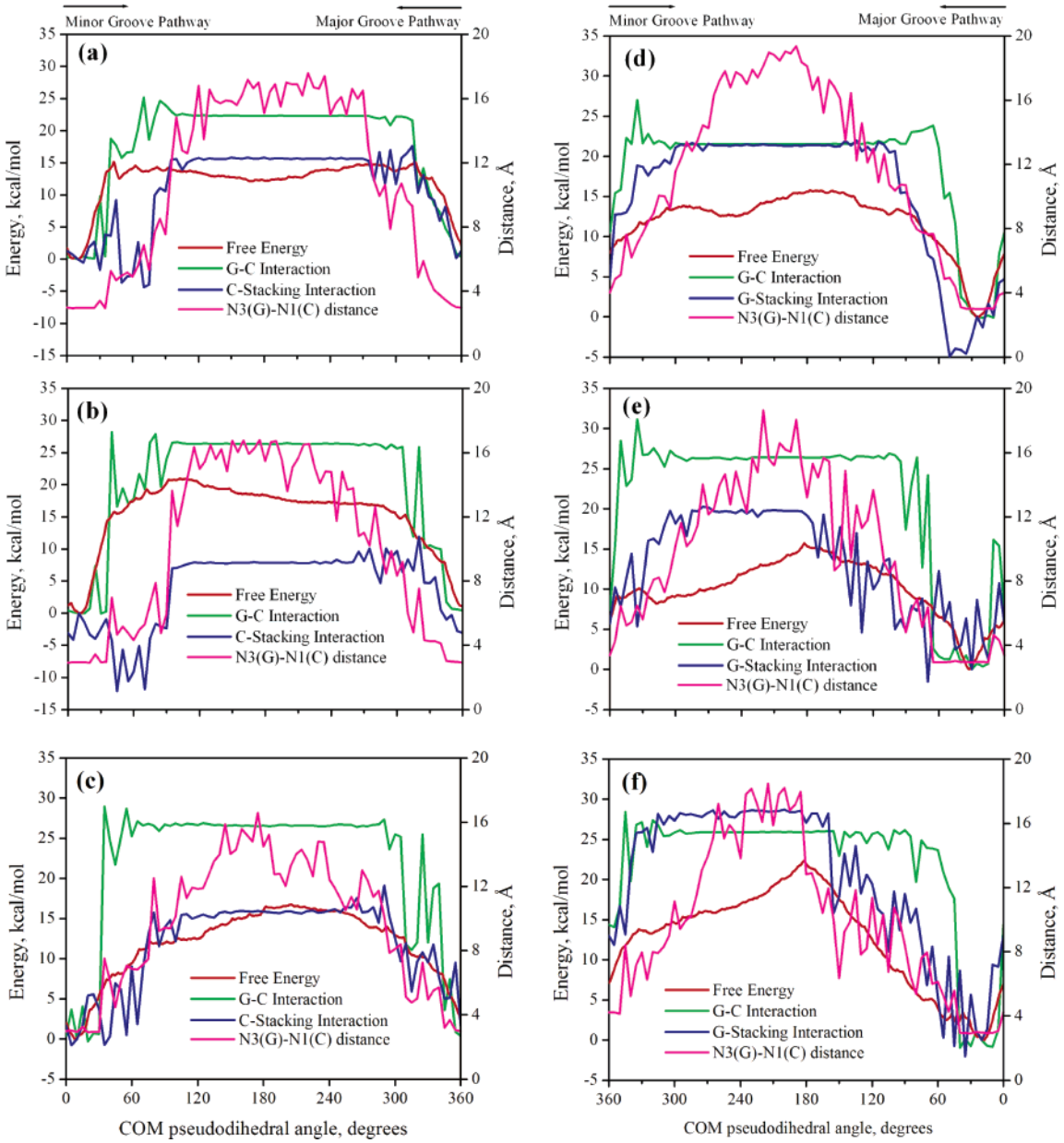


Figure 5. The change in the free energy (red), average N3(G)−N1(C) distance (purple), average G−C base pairing interaction energy (green), and average stacking interaction energy (blue) for C- (a, b, c) and G-base flipping (d, e, f) as a function of the COM pseudodihedral angle calculated using CHARMM (a and d), AMBER (b and e), and BMS (c and f). Base pairing interaction energies and stacking interactions were offset by the corresponding values at the WC base paired state. Stacking interaction energies were calculated with the neighboring bases of the same strand as the flipping base.

than the WC interaction energy. This is, in part, due to interactions of the flipping C base with the adjacent bases in the minor groove. These types of interactions have previously been reported and have been suggested to represent a mode for the effect of sequence on base flipping.[36] The fact that this phenomenon is observed in all three force fields indicates that it is not force field specific. With major groove C-flipping, the change in the stacking interaction energy is gradual, similar to that observed in the base pairing energies. The differences between the energetic changes during minor versus major groove flipping have previously been attributed to the need for the partner base to be "pushed" out of the way during minor groove flipping, which during major groove flipping gradually pulls away from the partner base.[36] Accordingly, the gradual loss of stacking energy is due to this type of motion during major groove flipping.

Stacking interactions during G-flipping show somewhat contrasting behavior to C-flipping (Figure 5). The AMBER and CHARMM stacking interaction energies are similar, while that with BMS is significantly more favorable. Here, the maintenance of stacking interactions while WC interactions are lost occurs via the major groove, with CHARMM showing a gain in stacking interactions in the vicinity of 50°. Consistent with the explanation for minor groove C-flipping this is due to the need for the flipping G base to push the partner C base out of its path during flipping. Again, interactions between the flipping base and the atoms in the grooves of the surrounding bases are observed.

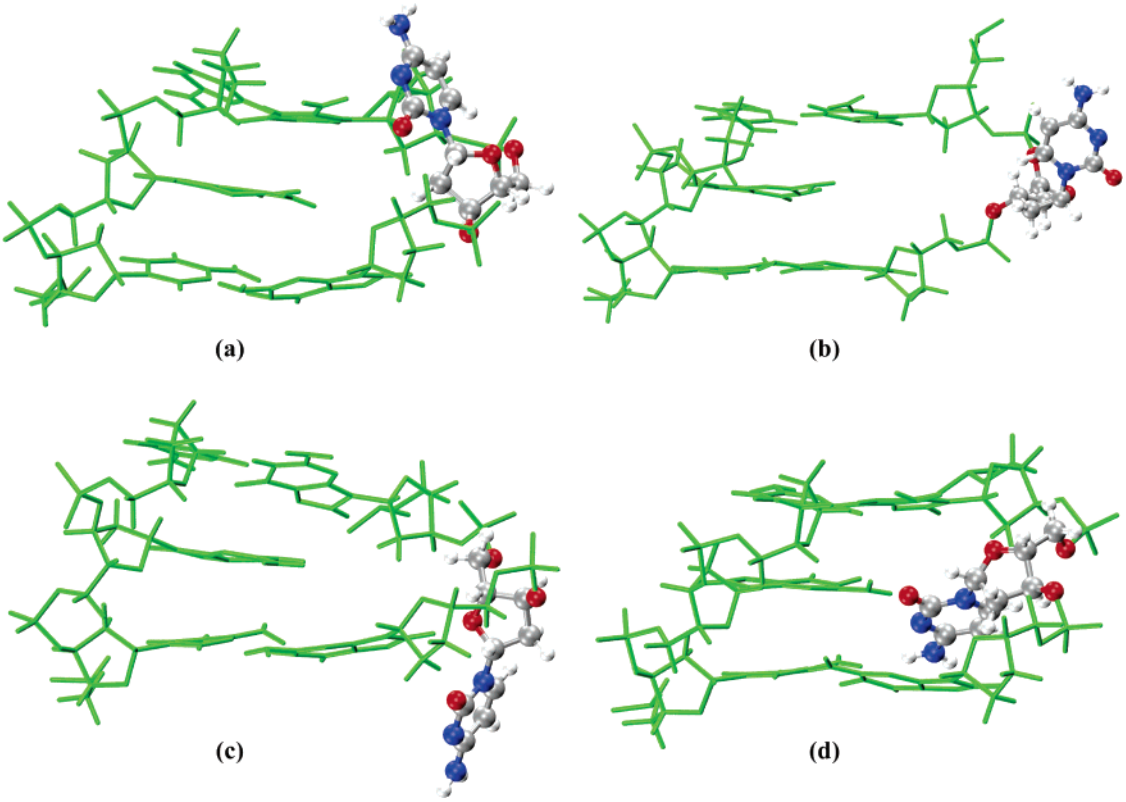


Figure 6. Representative structural snapshots from the MD simulations for C-flipping using the CHARMM force field for COM pseudodihedral angles of 80° (a), 115° (b), 280° (c), and 335° (d) showing the flipping base having favorable interactions with the adjacent bases or the backbone. Only the central tribase (GCG) is shown for clarity; the flipping base along with the sugar to which it is connected is shown in ball-and-stick representation.

Quantifying the accuracy of the force fields in their treatment of stacking is difficult. Previous comparisons of AMBER and CHARMM force fields with QM data show reasonable agreement with both models, with AMBER being in better agreement while CHARMM overestimates the interaction energies.[70] However, the quality of the QM data, which used the MP2/6-31G(d′) level of theory, where d′ indicates the use of a smaller exponent on the polarization function than in the normal basis set, is questionable. While the ordering of different stacking pairs and orientations may be reasonable from this level of theory due to the role of electrostatic interactions in that ordering, the absolute values may be less accurate due to limitations in that level of theory as well as QM methods in general in treating dispersion type interactions that dominate the total stacking interaction energy.[70−72] QM calculations with explicit consideration of electron correlation such as CCSD(T) accompanied with large basis sets are expected to provide more reliable data on the stacking interactions. However, state of art computational resources limit such calculations. In addition, the lack of a well-defined minima for these stacking complexes is another limitation.[48,71] Alternatively, is the validation of the stacking interactions via the calculation of experimental heats of sublimation, ΔH_{sub}, of base crystals. Such calculations have been performed for uracil and 9-methyladenine using the CHARMM force field,[13] showing good agreement with experimental data. Assuming that the in plane hydrogen bonding is being accurately treated by the CHARMM force field, as evidenced by the reproduction of experimental and QM data, the ability of the force field to reproduce the experimental ΔH_{sub} is a strong indication of its ability to accurately treat stacking. While the calculations with CHARMM have not been performed directly on the G and C bases, due to the absence of experimental data, its ability to reproduce the available experimental data suggests the model to be reasonably accurate with respect to the experimental regimen.

Solvent Contributions. Solvation effects have a strong impact on the properties of DNA, with one of the best examples being the change from the canonical B to A form of DNA as a function of decreasing water activity.[73] In the case of base flipping the movement of the base from the central region of the DNA duplex out of the helix leads to an increase in the exposure of the base to the aqueous environment. Solvation effects may be assessed by calculating solute−solvent interactions to assess the enthalpic contribution and by solvent accessible surface area (SASA), which accounts for the entropic contributions.[61] The flipping base is the one that experiences the most diverse solvation effects during the base flipping event. Also, the backbone of the DNA vicinal to the flipping base undergoes major conformational change and experiences varied exposure to the surroundings. Accordingly, change in the solvation energies and SASA of the flipping base and the backbone of the central tribase as a function of base flipping were analyzed. Solvation energy and SASA differences between the WC base paired and the fully flipped states are given in Table 2, with the average values for the individual states in

the Supporting Information Table S4. The solvation effects calculated undergo high fluctuations as a function of the pseudodihedral angle, hence, for the flipped state the means were obtained over windows, $x = 180-210°$. The interaction energies with solvent and SASA for all the windows as of function of the pseudodihedral are provided in the Supporting Information (Figures S3−S8). Expectedly, the solvation energies typically become more favorable, and the solvent accessibilities increase when the base flips out of the DNA duplex.

For the base alone, the change in the solvation energies of the C base in flipped DNA is predicted to be less favorable than that of the G base by all three force fields, consistent with the larger size and number of polar moieties on the G base. The differences in the solvation energies for the bases are similar with CHARMM and AMBER, while the BMS values are more favorable, as discussed above. Interestingly, while the change in SASA for the flipping bases are similar for CHARMM and AMBER (∼20 Å^2), they are significantly smaller for BMS, even though the solvation energies are significantly more favorable with the latter. This indicates that the base in BMS interacts more favorably with the solvent as compared to CHARMM and AMBER, while the smaller increase in SASA is due to enhanced interactions of the flipping base with the remainder of the DNA in BMS. Visual inspection of the final structures from the $x = 180°$ windows confirms this model (not shown).

Solvation analysis of the backbone of the flipping base shows relatively small changes upon flipping when considering the magnitude of the solvation energies and SASA values in the WC states (∼500 kcal/mol and ∼650 Å^2, respectively, Table S4, Supporting Information). In some cases the changes are slightly favorable, with the largest being −26 kcal/mol, with others close to zero. For the SASA, interestingly, in many cases there is a decrease in the accessibility in the flipped state. This appears to be due to the interactions of the flipping base with the local backbone atoms. Overall, these results indicate that changes in the solvation of the backbone are not significantly impacting the flipping PMFs.

Changes in solvation of the entire central tribase surrounding the flipping region, which includes both strands, were analyzed as changes would include contributions from the orphan bases and of the bases adjacent to the flipping base. For the tribase, in all cases the energies of solvation become more favorable in the flipped state, while the SASAs were larger. With CHARMM and AMBER, both the energy and SASA differences were similar for the two force fields as well as for C- versus G-flipping. However, with BMS, the magnitudes of the changes in both the energies of solvation and the accessibilities were smaller than with CHARMM and AMBER, though the direction of the change was the same.

Overall, the solvation results indicate only subtle differences between the three force fields with respect to flipping. Considering the base alone, BMS has more favorable solvation energies in the flipped states as compared to CHARMM and AMBER. This more favorable solvation would favor the flipped state, thereby contributing to the significantly larger equilibrium constant the open versus closed states for BMS (Table 1). However, analysis of the solvation of the central tribase shows CHARMM and AMBER to become more favorably solvated in the flipped states as compared to BMS. Thus, the present analysis does not allow for clear conclusions on the role of solvation on the calculated PMFs for the differenct force fields to be obtained.

Table 3: Differences in the Average Intrinsic Potential Energies between the WC and Fully Flipped States for Selected Regions of the DNA Using the CHARMM, AMBER, and BMS Force Fields[a]

	CHARMM	AMBER	BMS
	C-Flip		
tribase	54.7	54.3	38.9
backbone	12.6	7.7	−0.3
six bases	45.5	44.3	19.9
sugar	9.4	8.6	−1.5
phosphate	−0.3	−1.2	−0.8
flipping base	0.7	0.6	1.0
	G-Flip		
tribase	59.4	48.6	25.3
backbone	8.2	2.8	−0.9
six bases	44.9	45.2	37.0
sugar	8.4	2.0	−1.5
phosphate	0.4	−0.4	0.3
flipping base	0.4	−0.1	0.5

[a] All values are given in kcal/mol. Regions of the DNA include the central tribase and the corresponding backbone, six bases, sugar moieties, phosphate groups, and the flipping base. Individual energies along with error estimates from which the differences in this table were calculated are presented in Table S5 of the Supporting Information.

Intrinsic Potential Energy. Variation of the intrinsic potential energy of the DNA along the flipping pathway might be a factor affecting the free energy profiles. The difference of the intrinsic potential energies (i.e. internal molecular mechanical energy of the selected regions, including nonbond contributions) between the WC base paired and the flipped states obtained using the three force fields are given in Table 3. The values for the flipped states were taken as the average value of the windows with $x = 180-210°$. The relative values of the intrinsic potential energies with respect to the WC base paired state are given in Figure 7. Inspection of the figure quickly reveals that BMS yields the smallest increases in the intrinsic energies upon flipping, with the contribution in some cases being favorable. The potential energies of the phosphate groups and the flipping base do not vary much as predicted by all the three force fields. The increase in energy of the tribase, backbone, bases, and the sugar when the base flips out of the DNA duplex is quite substantial and are similar for CHARMM and AMBER. This increase in the energy due to the drastic conformational change seems to be significantly underestimated by BMS. It is interesting to note that the underestimation of the energies mainly corresponds to the neighboring bases of the target base as evidenced by the tribase and six bases results, and this is associated, in part, with the loss of WC and

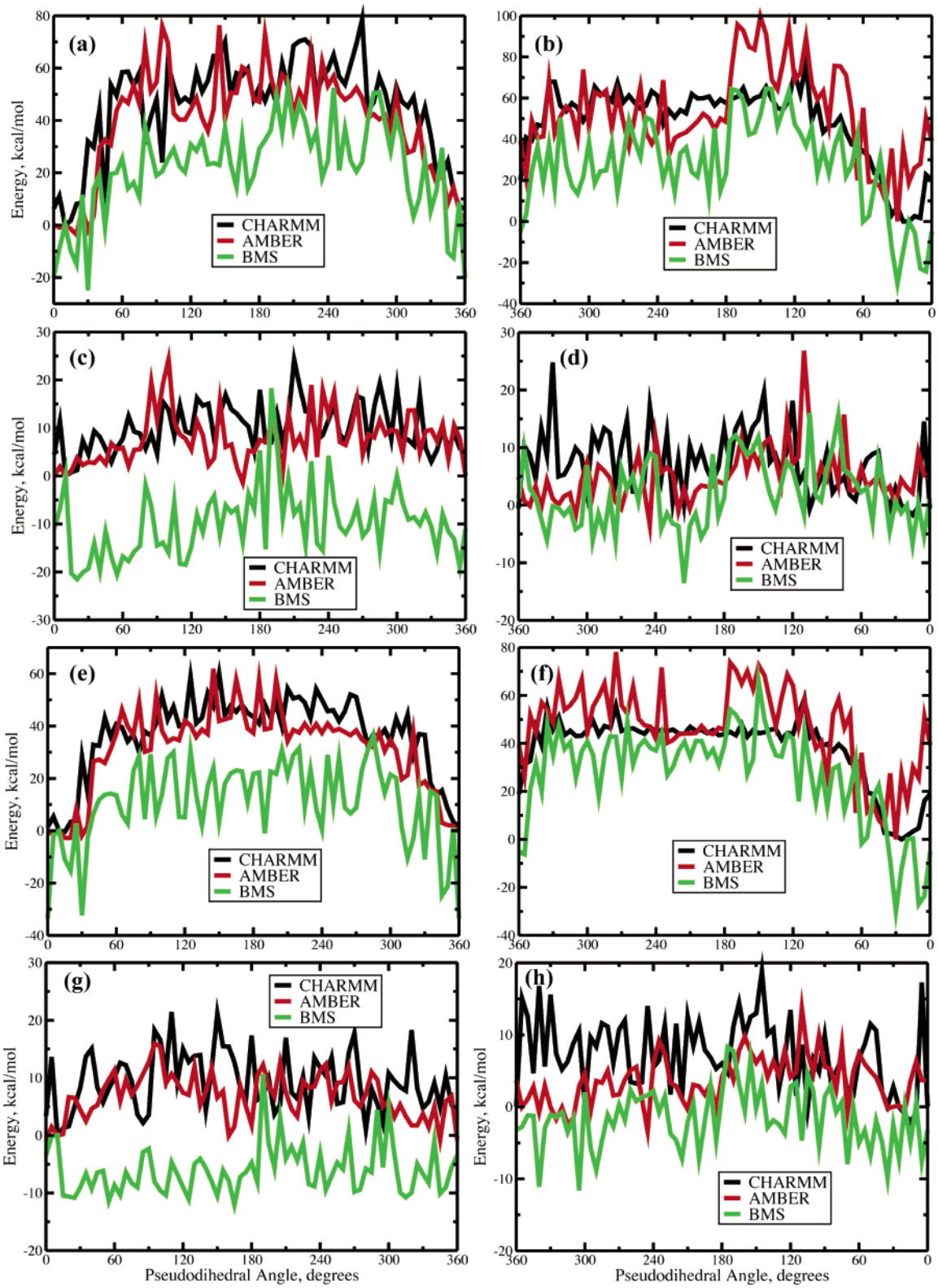


Figure 7. The change in the intrinsic potential energies of central tribase (a and b), backbone of the tribase (c and d), the six bases of the central tribase (e and f), and the sugar moieties of the tribase (g and h) obtained using CHARMM (black), AMBER (red), and BMS (green) corresponding to C-flipping (a, c, e, and g) and G-flipping (b, d, f, and h). The values are offset to those obtained for the WC base paired state.

stacking interactions, as discussed above. Thus, it appears that the intrinsic energetic contributions of the BMS force field leads to the preference of that model for the open state as predicted by the equilibrium constants (Table 1).

Conclusions

The present study reports a comparative assessment of the performance of three popular force fields for nucleic acids, CHARMM, AMBER, and BMS, to reproduce experimental

data from NMR imino proton exchange on the equilibrium between the base open and closed states associated with base flipping from DNA. Free energy profiles, or PMFs, were generated based on the umbrella sampling approach using a COM pseudodihedral restraint as the reaction coordinate. Convergence of the PMFs was critically analyzed; the results indicate that the amount of sampling via the 220 ps of sampling in each window of the PMF via MD simulations is satisfactory, and the conclusions arrived at from the present results are unlikely to change when additional sampling is performed. Comparison of the equilibrium constants between the open and closed states with the experimental data show the CHARMM27 force field to be in the best agreement, closely followed by AMBER with BMS being in significant disagreement. The tendency for both AMBER and BMS is to favor the open state. In addition, CHARMM is consistent with the experimental observation that the base flipped state corresponds to a metastable state lying around 12 kcal/mol above the WC base paired state. However, it should be emphasized that the present results are limited to a single base pair in a single sequence, such that the generality of the present observations requires additional studies.

Qualitatively, the free energy profiles calculated using the three force fields differ significantly with respect to the shape of the surfaces including barrier heights and the presence and depth of stable minima associated with the flipped states. Various components, namely the base pairing and stacking interaction energies and solvation energies, assumed to contribute to the energetics of base flipping, were assessed and shown to correlate with the observed free energy change. The base pair interaction energies for WC GC basepair calculated using the AMBER and BMS force fields are more favorable as compared to CHARMM, with the present values from AMBER and BMS being in disagreement with QM and experimental data. With base stacking significant variations between the force fields are present, though it is currently difficult to evaluate the accuracy of the force fields based on QM data. Changes in solvation energies of the flipping base also differ significantly between the three force fields with CHARMM and AMBER being more similar, while BMS shows more favorable solvation of the flipping base. However, the change in solvation of the central tribase is less favorable with BMS than with AMBER and CHARMM, making it difficult to draw conclusions concerning the solvation contributions to the free energy profiles. Finally, analysis of the intrinsic potential energies of the DNA as a function of flipping indicate systematic differences; CHARMM and AMBER are similar and unfavorable for the flipped states, while those terms with BMS are significantly less unfavorable and, in some cases, slightly favorable for the flipped state. These results indicate that the favoring of the open state by BMS is dominated by intrinsic energetic contributions.

Overall, the results speak to the quality of both CHARMM and AMBER in modeling the structural distortion of DNA associated with base flipping. On the other hand, BMS favors the more open state, although previous studies have shown this force field to model canonical crystal structures of B-form DNA.[7,9,10] However, the individual contributions from the different force fields to the flipping PMFs in some cases vary significantly. Such differences indicate that the behavior of the force fields are based, to some extent, on different relative contributions for different parts of the model and, importantly, emphasize that such force field effects must be taken into account when interpreting results from MD simulations.

Acknowledgment. We acknowledge NIH Grant GM51-501 for financial support and the Pittsburgh Supercomputing Center for computational support. We express appreciation to Dr. Irina Russu for helpful discussions.

Supporting Information Available: Figures depicting the free energy profiles calculated using various windows, target base−DNA interaction, solute−solvent interaction, and solvent accessibilities and tables presenting the equilibrium constants calculated using various ranges of windows, and solvation energies, solvent accessible surface area, and intrinsic potential energies of the WC base paired and flipped states. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Karplus, M.; McCammon, J. A. *Nature Sturct. Biol.* **2002**, *9*, 646−652.

(2) Karplus, M. *Acc. Chem. Res.* **2002**, *35*, 321−323.

(3) *Computational Biochemistry and Biophysics*; Becker, O. M., MacKerell, A. D., Jr.; Roux, B., Watanabe, M., Eds.; Marcel-Dekker: New York, 2001; p 512.

(4) MacKerell, A. D., Jr. *J. Comput. Chem.* **2004**, *25*, 1584−1604.

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

(6) Auffinger, P.; Westhof, E. *Curr. Opin. Struct. Biol.* **1998**, *8*, 227−236.

(7) Cheatham, T. E., III. *Curr. Opin. Struct. Biol.* **2004**, *14*, 360−367.

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

(9) Cheatham, T. E., III.; Young, M. A. *Biopolymers* **2000**, *56*, 232−256.

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

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

(12) MacKerell, A. D., Jr.; Nilsson, L. Nucleic Acid Simulations. In *Computational Biochemistry and Biophysics*; Becker, O. M., MacKerell, A. D., Jr., Roux, B., Watanabe, M., Eds.; Marcel Dekker: New York, 2001; pp 441−464.

(13) Foloppe, N.; MacKerell, A. D., Jr. *J. Comput. Chem.* **2000**, *21*, 86−104.

(14) MacKerell, A. D., Jr.; Banavali, N. K. *J. Comput. Chem.* **2000**, *21*, 105−120.

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

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

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

(18) Feig, M.; Pettitt, B. M. *J. Phys. Chem. B* **1997**, *101*, 7361−7363.

(19) Feig, M.; Pettitt, B. M. *Biophys. J.* **1998**, *75*, 134−149.

(20) Cheatham, T. E., III.; Kollman, P. A. *J. Mol. Biol.* **1996**, *259*, 434−444.

(21) Reddy, S. Y.; LeClerc, F.; Karplus, M. *Biophy. J.* **2003**, *84*, 1421−1449.

(22) Cheng, X.; Roberts, R. J. *Nucleic Acids Res.* **2001**, *29*, 3784−3795.

(23) Hornby, D. P.; Ford, G. C. *Curr. Opin. Biotechnol.* **1998**, *9*, 354−358.

(24) Roberts, R. J. *Cell* **1995**, *82*, 9−12.

(25) Roberts, R. J.; Cheng, X. *Annu. Rev. Biochem.* **1998**, *67*, 181−198.

(26) Stivers, J. T. *Prog. Nucleic Acid Res. Mol. Biol.* **2004**, *77*, 37−65.

(27) Goedecke, K.; Pignot, M.; Goody, R. S.; Scheidig, A. J.; Weinhold, E. *Nature Struct. Biol.* **2001**, *8*, 101−103.

(28) Klimasauskas, S.; Kumar, S.; Roberts, R. J.; Cheng, X. *Cell* **1994**, *76*, 357−369.

(29) Cheng, X.; Blumenthal, R. M. *Structure* **1996**, *4*, 639−645.

(30) Lyakhov, I. G.; Hengen, P. N.; Rubens, D.; Schneider, T. D. *Nucleic Acids Res.* **2001**, *29*, 4892−4900.

(31) Varnai, P.; Canalia, M.; Leroy, J. L. *J. Am. Chem. Soc.* **2004**, *126*, 14659−14667.

(32) Dornberger, U.; Leijon, M.; Fritzsche, H. *J. Biol. Chem.* **1999**, *274*, 6957−6962.

(33) Gueron, M.; Leroy, J. L. *Methods Enzymol.* **1995**, *261*, 383−413.

(34) Chen, C. J.; Russu, I. M. *Biophys. J.* **2004**, *87*, 2545−2551.

(35) Moe, J. G.; Foltastogniew, E.; Russu, I. M. *Nucleic Acids Res.* **1995**, *23*, 1984−1989.

(36) Banavali, N. K.; MacKerell, A. D., Jr. *J. Mol. Biol.* **2002**, *319*, 141−160.

(37) MacKerell, A. D., Jr.; Brooks, B.; Brooks, C. L., III.; Nilsson, L.; Roux, B.; Won, Y.; Karplus, M. CHARMM: The Energy Function and Its Paramerization with an Overview of the Program. 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, 1998; Vol. 1, pp 271−277.

(38) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.

(39) Quanta. 4.0 ed.; Accelrys Inc.: San Diego, CA, 2001.

(40) Field, M. J.; Karplus, M. CRYSTAL Module of CHARMM, 22nd ed.; Harvard University: Cambridge, MA, 1992.

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

(42) Steinbach, P. J.; Brooks, B. R. *J. Comput. Chem.* **1994**, *15*, 667−683.

(43) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327−341.

(44) Nosé, S. *J. Chem. Phys.* **1984**, *81*, 511−519.

(45) Beglov, D.; Roux, B. *J. Phys. Chem. B* **1997**, *101*, 7821−7826.

(46) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011.

(47) Crouzy, S.; Baudry, J.; Smith, J. C.; Roux, B. *J. Comput. Chem.* **1999**, *20*, 1644−1658.

(48) Pan, Y. P.; Priyakumar, U. D.; MacKerell, A. D. *Biochemistry* **2005**, *44*, 1433−1443.

(49) Huang, N.; Banavali, N. K.; MacKerell, A. D., Jr. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 68−73.

(50) Giudice, E.; Várnai, P.; Lavery, R. *ChemPhysChem* **2001**, *11*, 673−677.

(51) Giudice, E.; Lavery, R. *J. Am. Chem. Soc.* **2003**, *125*, 4998−4999.

(52) Giudice, E.; Varnai, P.; Lavery, R. *Nucleic Acids Res.* **2003**, *31*, 2703−2703.

(53) Varnai, P.; Lavery, R. *J. Am. Chem. Soc.* **2002**, *124*, 7272−7273.

(54) Mol, C. D.; Izumi, T.; Mitra, S.; Tainer, J. *Nature* **2000**, *403*, 451−456.

(55) Wärmländer, S.; Sen, A.; Leijon, M. *Biochemistry* **2000**, *39*, 607−615.

(56) Warmlander, S.; Sandstrom, K.; Leijon, M.; Graslund, A. *Biochemistry* **2003**, *42*, 12589−12595.

(57) Guéron, M.; Kochoyan, M.; Leroy, J.-L. *Nature* **1987**, *328*, 89−92.

(58) Klimasauskas, S.; Szyperski, T.; Serva, S.; Wuthrich, K. *EMBO J.* **1998**, *17*, 317−324.

(59) Truhlar, D. G.; Garrett, B. C.; Klippenstein, S. J. *J. Phys. Chem.* **1996**, *100*, 12771−12800.

(60) Albery, W. J. *Adv. Phys. Org. Chem.* **1993**, *28*, 139−170.

(61) Lee, B.; Richards, F. M. *J. Mol. Biol.* **1971**, *55*, 379−400.

(62) Brameld, K.; Dasgupta, S.; Goddard, W. A., III. *J. Phys. Chem. B* **1997**, *101*, 4851−4859.

(63) Hobza, P.; Kabelac, M.; Sponer, J.; Mejzlik, P.; Vondrasek, J. *J. Comput. Chem.* **1997**, *18*, 1136−1150.

(64) Yanson, I. K.; Teplitsky, A. B.; Sukhodub, L. F. *Biopolymers* **1979**, *18*, 1149−1170.

(65) Cheng, Y. K.; Pettitt, B. M. *Prog. Biophys. Mol. Biol.* **1992**, *58*, 225−257.

(66) Sponer, J.; Leszczynski, J.; Hobza, P. *Biopolymers* **2001**, *61*, 3−31.

(67) Kool, E. T. *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 1−22.

(68) Gago, F. *Methods-a Companion Methods Enzymol.* **1998**, *14*, 277−292.

(69) Hobza, P.; Sponer, J. *J. Am. Chem. Soc.* **2002**, *124*, 11802−11808.

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

(71) Hobza, P.; Selzle, H. L.; Schlag, E. W. *J. Phys. Chem.* **1996**, *100*, 18790−18794.

(72) Hobza, P.; Selzle, H. L.; Schlag, E. W. *J. Am. Chem. Soc.* **1994**, *116*, 3500−3506.

(73) Saenger, W.; Hunter, W. N.; Kennard, O. *Nature* **1986**, *324*, 385−388.

CT0501957

JCTC Journal of Chemical Theory and Computation

Comparison of Protocols for Calculation of Peptide Structures from Experimental NMR Data

Marc Fuhrmans,[†] Alexander G. Milbradt,[†] and Christian Renner*[,†,‡]

Max-Planck-Institut für Biochemie, Martinsried, Germany, and School of Biomolecular and Natural Sciences, Nottingham Trent University, Clifton Lane, Nottingham NG11 8NS, United Kingdom

Received August 13, 2005

Abstract: In a comparison of structure calculation protocols we clearly demonstrate the need for generating independent starting structures, which is for peptides most efficiently achieved by distance geometry (DG) methods. Our test set consisted of 20 peptides with 7−9 amino acid residues additionally constrained by backbone cyclization and/or the presence of a disulfide bridge. Small peptides usually adopt defined conformational properties only upon introduction of additional constraints, such as cyclization. Therefore, we believe the results of our comparison to be applicable to a large and important class of molecules. The problems associated with the use of restrained molecular dynamics (MD) for conformational searching in the context of structure calculation consist in energy barriers that derive mainly but not exclusively from the experimental NOE constraints. A valid alternative to the DG approach, although for peptides computationally less efficient, is MD simulated annealing starting from random structures as commonly performed in the protein structure calculation from NMR data. As a consequence of our study it must be expected that a considerable fraction of published peptide structures are artificially well-defined or even wrong. Given the relevance of peptide studies for both drug development and protein folding we regard it highly important that structure calculations of peptides are performed with more consideration.

Introduction

In living organisms the blueprints for building proteins are stored in the form of amino acid sequences on genes. The translation of this primary structure into a folded and functional protein is effected by the inherent properties of the amino acids and the cellular environment including chaperones, folding adjuvants, and proper folding conditions. In this highly efficient way only a small amount of information needs to be stored (primary sequence) for generating macromolecules with oftentimes very complex structural and dynamic properties. Unfortunately, the translation rules are exceedingly complicated making protein folding one of the most challenging problems of biochemistry. For these reasons considerable experimental effort needs to be invested for determining three-dimensional structures of proteins although the primary sequence is already available or can be determined in a straightforward manner for a given gene of interest. Whereas X-ray crystallography dominates the structure determination of proteins,[1] for small peptides NMR spectroscopy is the method of choice because most peptides do not crystallize. Technically, peptide NMR is less complicated than the highly sophisticated multidimensional heteronuclear experiments used in protein NMR.[2] The differences in approaching small peptides or proteins by NMR are easily understood as peptides comprise a much smaller number of resonances, and therefore problems of overlap are of lesser concern. However, due to the generally higher flexibility of peptides and the smaller number of experimental constraints obtainable from the NMR spectra, conversion of NMR data to three-dimensional

* Corresponding author phone: +44-115-848-3522; fax: +44-115-848-6636; e-mail: christian.renner@ntu.ac.uk.

[†] Max-Planck-Institut für Biochemie.

[‡] Nottingham Trent University.

10.1021/ct050203j CCC: $33.50
Published on Web 12/06/2005

structures, i.e., structure calculation, is of higher concern for peptides.[3−10] For well-structured proteins the conformational space available to the molecule is so restricted as to leave less room for variations depending on the details of the structure calculation protocol. Contrarily, in peptides often multiple conformations occur, and their correct representation in the final structural ensemble might be sensitive to the calculation strategy.[4−6,10]

The two most widely used approaches to NMR conformational analysis of peptides are distance geometry (DG) based and pure molecular dynamics (MD) protocols, both used since the beginning of peptide NMR. For calculation of three-dimensional structures sufficient sampling of the conformational space is possibly the most important factor.[10] While MD simulations should in principle detect any possible conformation of the molecule according to the ergodicity theorem, the required time might exceed the current computational limits by far. Alternatively, more direct methods for sampling the conformational space can be used: different DG methods[11−16] and related approaches, e.g. refs 17−19, have been devised for structure calculation based on experimental distance constraints as obtained by NMR. Again good sampling properties are vital for the performance of these strategies.[14,15,20−22] Despite the existing discussion in the literature about conformational sampling, pure MD protocols are frequently used in published NMR studies of peptides. The applicability of the specific protocol and the possibility of incomplete sampling of the conformational space are usually not addressed. Comparisons of structure calculation methods have been reported before,[3,13,23−25] but only for one or two molecules in each case. We want to demonstrate the shortcomings of the simple MD method on a larger set of molecules. Furthermore, we compare for our set of peptides the results from the typical peptide protocols to those obtained with the structure calculation protocols that are commonly used for proteins.

Methods

General. Distance geometry[12] and molecular dynamics-simulated annealing (MD-SA)[26] calculations were performed with the INSIGHTII (version 2000) software package (Accelrys, San Diego, CA) on Silicon Graphics O2 R5000 computers (SGI, Mountain View, CA). In each calculation 100 structures were generated either by distance geometry, by assigning random values to the coordinates, or by a molecular dynamics run of 1 ns, where one structure was saved each 10 ps. In all cases the 100 structures were refined with a short MD-SA protocol: After an initial minimization, 5 ps at 300 K were simulated followed by exponential cooling to ~0 K during 10 ps. The refinement for the random structures included an additional 2 ps at 500 K prior to the 5 ps at 300 K. The cooling phase was reduced to 8 ps in this case resulting in the same overall length of 15 ps for the refinement step. This modification was introduced, because releasing residual strain was found to be more difficult for the structures, which were derived from random coordinates. The final structures were sorted according to their final energy, and the 20 energy-lowest were analyzed. A time step of 1 fs was used with the CVFF force field[27] while simulating the solvents DMSO and H_2O with dielectric constants of 46.7 and 80.0, respectively. For some examples additional calculations were performed in an identical manner but using the AMBER force field.[28] The experimental constraints were applied at every stage of the calculations with the same force constants as for the published structures.[29−32]

Distance Geometry protocol ("DG/MD"). One hundred structures were generated from distance-bound matrices.[12] Triangle-bound smoothing and prospective metrization were used. The structures were generated in four dimensions, then reduced to three dimensions, and optimized with a simulated annealing step according to the standard protocol of the DG II package of INSIGHT II. DG calculations generally result in poor covalent geometry (bond length, angles, etc). Therefore, in addition to the coarse optimization of the standard protocol a subsequent MD-SA refinement with DISCOVER was performed (see section *General* above).

Simple MD Protocol ("Pure MD"). All molecular dynamics calculations were performed with the DISCOVER module of INSIGHTII. As the starting point for the generation of 100 structures the energy-lowest structure of the published NMR ensemble[29−32] was used for each molecule. Velocities for this starting structure were generated at 10 K, and the system was then heated to 1000 K during 50 ps (temperature bath coupling with a 5 ps time constant). During the following 1 ns production run at 1000 K one structure was saved each 10 ps for further refinement (see above). In many cases an additional structure calculation was performed with a second starting structure corresponding to a low energy structure of the published ensemble that was conformationally dissimilar to the first starting structure and was not sampled in the first MD run.

MD Protocol with Reduced Force Field during the Conformational Sampling ("Scaled MD"). The only difference to the previous protocol (pure MD) was that during the conformational sampling at 1000 K nonbonded interactions (van der Waals and Coulomb) were reduced to 10%, while the through-bond interactions were scaled down to 50%. For the annealing step the force field was applied with full strength.

MD Protocol Starting from Random Coordinates ("Random MD"). Starting from random coordinates the first step consisted in achieving approximately reasonable covalent geometry: van der Waals and electrostatic interactions were scaled down to 1%, while the force constants for bond lengths, angles, and dihedrals were only reduced to 50%. After minimization nonbonded interactions were scaled up to 10% (as for the scaled MD). Another minimization followed before 10 ps could be simulated at 1000 K. Finally, the force field was restored to its normal strength for the simulated annealing step.

Results

Multiple structure calculations were performed for 20 peptides (Figure 1) that have been investigated recently in our laboratory (refs 29−32 and unpublished results) according to different structure calculation protocols (Figure 2). Most of the 20 peptides are similar in that they contain a peptide

Cyclic backbone with *cis/trans* azo group

Disulfide bridge		No disulfide bridge	
c[APB-ACATCDGF]	(1 c/t)	c[AMPB-KARGDfV]	(7 c/t)
c[AMPB-ACATCDGF]	(2 c/t)	c[APB-ACATCDGF]	(8 c/t)
c[AMPB-KCATCDKK]	(3 c/t)	StBu StBu	
c[AMPB-KCGHCKKK]	(4 c/t)	c[AMPB-ACATCDGF]	(9 c/t)
c[APB-ACATCDGFF]	(5 t)	StBu StBu	
c[AMPB-KCATCDKKK]	(6 t)	c[AMPB-KSATSDKK]	(10 c/t)

Linear

Disulfide bridge		No disulfide bridge	
ACATCDGF	(11)	GWGQPHGG	(12)

Figure 1. Peptides for which NMR data were determined previously in our group. Azobenzene containing peptides (**1**−**10**) can occur in two isomeric forms cis and trans of the azo moiety indicated by (c/t). Whereas trans is the ground state, also the cis isomer has a lifetime long enough to perform NMR experiments. Because the geometry of trans and cis azobenzene is completely different, the two isomers are treated as separate molecules (e.g. **1 cis** and **1 trans**) for the purpose of the present work. For **5** and **6** the structures were only determined for the trans isomer. Note: f = d-Phe, APB = 4-(amino)phenylazobenzoic acid, AMPB = 4-(aminomethyl)-phenylazobenzoic acid, tBu = *tert*-butyl.

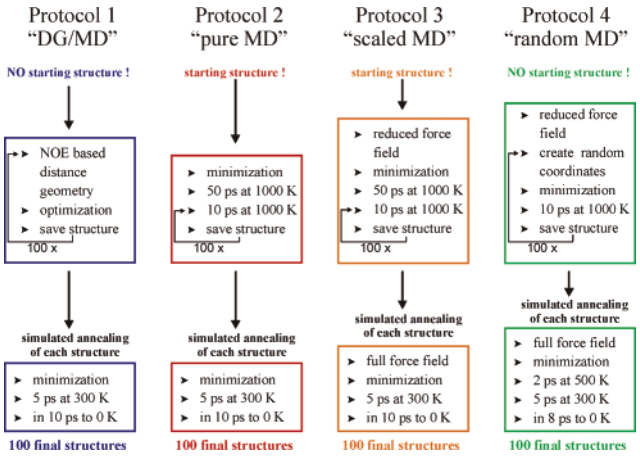


Figure 2. Flowcharts for the different protocols that were used for calculating structural ensembles using NMR data. The final structures were sorted according to energy and the 20 energy-lowest were analyzed.

stretch of 7−9 amino acid residues backbone-cyclized by 4-(amino)phenylazobenzoic acid or 4-(aminomethyl)phenylazobenzoic acid. This similarity allows for the systematic investigation of the influence of ring size and additional cyclization (i, i+3 disulfide bridge). For reference a peptide sequence with the i, i+3 disulfide bridge (**11**), but not backbone-cyclized, as well as a linear unconstrained octapeptide (**12**) were used. Figure 2 shows that each protocol consists of a first part for generating structures and a second part consisting of a simulated annealing step. The annealing step is identical for all protocols except for the fourth where a short 500 K phase was inserted before the simulation at 300 K. To achieve the same overall length the final cooling period is shortened correspondingly. In the first part the

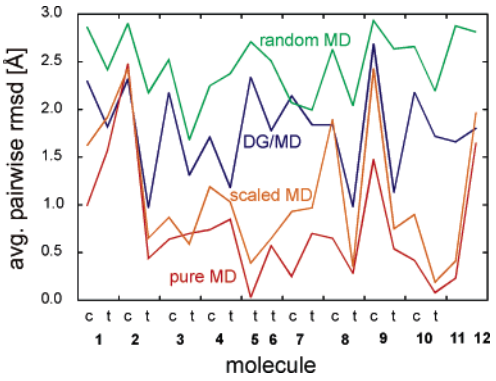


Figure 3. Conformational variability among the 20 energy-lowest structures obtained with different protocols for each molecule. Only backbone atoms were used for calculation of the rmsd.

structure generation consists of 10 ps at 1000 K per structure for protocols 2−4 that do not use distance geometry. In this way the computational costs for protocols 2−4 are nearly identical, and variations between protocols are kept to a minimum. Note that protocol 1 is the one that was used for generating the published structures.[29−32] Our protocols do not necessarily represent the most common or optimized versions of the respective approach. Instead they were chosen in such a fashion that they are as similar as possible to allow for interpretation of differences in the resulting structural ensembles. Specifically, we have placed an emphasis on exploration of conformational space represented by the first part of our protocols. Published MD protocols often employ multiple annealing cycles of a few picoseconds each for improved sampling and location of energy minima. While we used only one final annealing step for location of the energy minima, our conformational search comprised 1000 ps at 1000 K and thus clearly surpasses common MD strategies in terms of conformational sampling. With each protocol 100 structures were calculated for each of the 20 peptides. The 20 energy-lowest of each structural ensemble were analyzed with respect to NOE violations, final energy, and conformational variability as expressed by the average pairwise rmsd. The average pairwise rmsd is calculated as the average of the rmsd values of all possible pairs i and j with $i \neq j$ being two structures of the respective structure ensemble. For comparing two ensembles the pairwise rmsd can be calculated with pairs i and j, where i is from one ensemble and j is from the other. All pairwise rmsds can be displayed in so-called cluster graphs that are 3D diagrams where x and y define the number of the structures i and j and the corresponding rmsd(x, y) constitutes the third dimension. Analysis of cluster graphs calculated for two concatenated structural ensembles allows for detailed comparison of both. We consider two structures distinctly different, if their rmsd is greater than 2 Å. This limit was found useful for defining and differentiating conformational families. The average pairwise rmsd for one ensemble results in slightly larger values as the more familiar average rmsd to the average structure. However, the latter is not as easily generalized to a comparison of two ensembles.

Figure 3 compares the conformational homogeneity or heterogeneity of structural ensembles obtained using the

Table 1. Comparison of DG/MD and Pure MD Approach for 20 Peptides[c]

peptide [a]	homogeneous ensemble			DG ≈ MD	MD ≠ MD (1st) (2nd) energy barrier
	DG	MD	MD [b] (2nd)		
1 *cis*		X	X		
1 *trans*				X	
2 *cis*				X	
2 *trans*	X	X		X	
3 *cis*		X	X		
3 *trans*	X	X		X	
4 *cis*		X	X		X
4 *trans*	X	X		X	
5 *trans*		X	X		X
6 *trans*		X	X		X
7 *cis*		X	X		X
7 *trans*		X	X		X
8 *cis*		X	X		X
8 *trans*	(X)	X	X	(X)	
9 *cis*					X
9 *trans*	X	X		X	
10 *cis*		X	X		
10 *trans*		X	X		
11		X	X		X
12	X	X		X	

[a] Color-coding of peptides: green: DG and MD result in similar ensembles; red: energy barrier in MD simulation; black: no energy barrier, but MD ensemble incomplete. [b] Second MD from a second starting structure that is contained in the DG ensemble but not sampled in the first MD calculation. [c] Peptides **1**−**6** (bicyclic) and peptides **7**−**10** (monocyclic) are sorted according to increasing ring size.

protocols of Figure 2 for all 20 peptides. Obviously, generation of structures by molecular dynamics results mostly in more homogeneous ensembles than the calculation of independent structures either by DG or random starting coordinates. This result is expected, because sampling of conformational space might be impeded in molecular dynamics by high energy barriers (protocols 2 and 3), while protocols 1 and 4 directly start from distinct points in conformational space and, thus, circumvent the problem of high energy barriers (but not that of low barriers and general roughness of the energy surface, as will be seen below). The fact that our DG ensembles capture a larger part of the accessible conformational space compared to the MD ensembles, however, does not prove or even indicate that sampling of the DGII method as implemented in the INSIGHT2 software is ideal or complete (see refs 20−22 for a comparison of various DG methods). A detailed comparison of the structural ensembles of DG/MD and pure MD reveals that in cases when the DG ensemble consists of only one conformational family the same family is also found with the pure MD protocol albeit with partially lower rmsds. However, when more than one conformational family is present in the DG ensemble, often the pure MD reproduces only one of them. In these cases an additional MD calculation (according to protocol 2) was performed starting from a conformation that was present in the DG ensemble but not sampled in the first MD run. For this purpose the DG structure was compared to all 100 structures of the first MD ensemble, not only the 20 energy-lowest. Table 1 summarizes the results of the additional MD calculations. For almost half of the peptides well defined, but dissimilar ensembles were obtained with the same calculation protocol (#2), documenting a strong dependence on the starting structure. In Table 1 peptides are sorted according to constraints imposed by cyclization with bicyclic peptides and small ring sizes at the top and less constrained peptides at the bottom. Inspection of the last column of Table 1 reveals the occurrence of an energy barrier with little correlation to ring size or presence of the additionally constraining disulfide bridge. Therefore, it seems not possible to anticipate the presence or absence of energy barriers in MD structure calculation based on the given chemical structure of the peptide. Although the presence of energy barriers should always be considered, the severity of this problem even at a simulation temperature of 1000 K might surprise those who have so far relied on a variation of the pure MD protocol. Figure 4 exemplifies that apparent structural similarity is also no clue to the presence of energy barriers: Four peptides are shown for which protocol #2 (pure MD) was performed with two starting structures that both belong to the published NMR structural ensemble. While the two starting structures are quite similar for peptides **7 cis** and **8 trans**, those of **7 trans** and **10 trans** exhibit pronounced differences. The resulting ensembles were homogeneous and reasonably well-defined (see Table 1, structures not shown). Structure calculations have to be independent of the starting structure so that the same final ensembles should be obtained for both starting structures. While this was indeed observed for **8 trans** and **10 trans**, shown on the right side in Figure 4, for **7 trans** and **7 cis** the result depended strongly on the starting structure. For the four peptides of Figure 4 obviously no correlation exists between the apparent similarity of the two starting structures and the presence of a considerable energy barrier. For peptide **10 trans** the second MD calculation resulted in basically the same ensemble as the first MD, although the starting structure was different. In these cases the question remains, whether the second starting structure that is contained in the DG, but not in the MD ensemble is a realistic conformation of the peptide or not. We think that every conformation that satisfies the experimental constraints and is compatible with the force field (i.e. low energy) has to be considered as realistic.

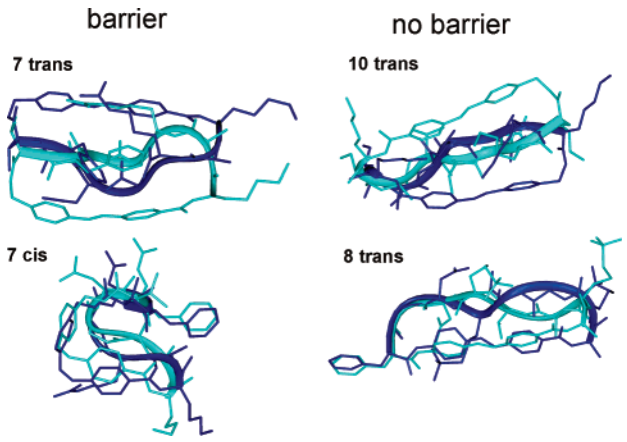


Figure 4. Comparison of the starting structures of two different MD calculations for selected peptides (**7 trans, 10 trans**: starting structures with pronounced difference; **7 cis, 8 trans**: similar starting structures). For peptides on the right very similar results were obtained from the MD simulations, whereas for peptides on the left results depended heavily on the starting structure pointing to the existence of an energy barrier.

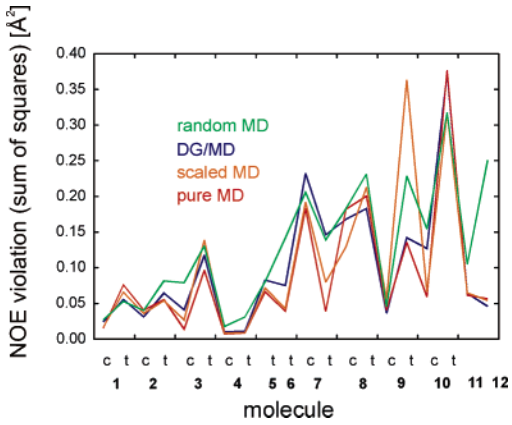


Figure 5. Squared NOE violations were summed up for each structure and averaged over the 20 energy-lowest structures. Results obtained with different protocols are compared for each molecule.

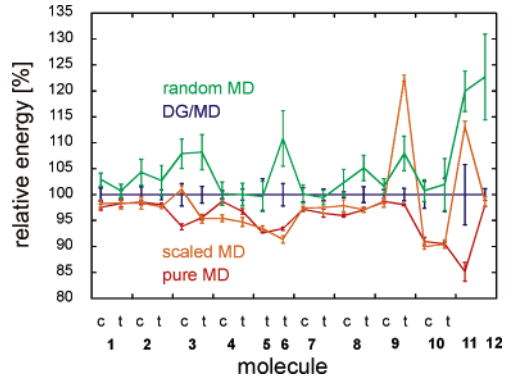


Figure 6. Average energy of the 20 energy-lowest structures relative to the average energy obtained with the DG/MD protocol. Error bars indicate the relative standard deviation, i.e., standard deviation divided by average.

NOE violations are compared in Figure 5. For this graph the squared NOE violations were summed up for each structure, and the resulting sums were averaged over the 20 energy-lowest structures of an ensemble. Multiplication of the sums of squared NOE violations with a force constant of 50 kcal $Å^{-2}$ would yield the corresponding energy penalty. It is seen from Figure 5 that experimental constraints were fulfilled in the different ensembles to similar extents. However, for the final energies clear trends are visible in Figure 6. As the absolute values of the energies have no real meaning, final energies relative to those obtained with protocol 1 are depicted. Clearly, protocol 4 (random MD) results in the most unfavorable energies, whereas the pure MD ensemble exhibits the smallest values. This comparison indicates that the structures of the random MD ensemble and, also to some extent, those of the DG ensemble are not fully relaxed. For the DG ensemble we tested in additional calculations that for most peptides more extensive simulated annealing can indeed reduce the final energies to values close to those of the corresponding pure MD ensemble (data not shown). While sufficiently long or iterated steps of simulated annealing should result in very similar final energies for all four protocols, we have purposely chosen annealing steps of only moderate length for comparing the efficiency of the first step of the protocols where structures are generated (DG vs MD vs random coordinates). We consider a structure generating method more efficient, if the resulting ensemble can be refined more easily to agree with the MD force field as indicated by lower energies. Extensive minimization either by steepest descent or conjugate gradient methods did not significantly reduce the energies of either the DG/MD or pure MD ensemble. Apparently small energy barriers and the general roughness of the energy landscape complicate localization of conformations with minimal energy during the annealing step. For some of the peptides it has even been shown experimentally by ultrafast time-resolved spectroscopy that their energy landscapes are surprisingly complex despite their small sizes.[33−36]

Discussion

Although NMR is the only technique to date for determining the structures of flexible or semiflexible peptides experimentally with atomic resolution, certain limitations inherent in this kind of spectroscopy have to be taken into account. In addition to aspects of correct identification and quantification of geometric constraints[37] the relation between the amount of experimental data and conformational heterogeneity is of concern. The lack of NOE constraints is a consequence of conformational flexibility[38] However, spectral overlap and special amino acid sequences can also prevent extraction of sufficient NOEs for an unambiguous well-defined structure (e.g. poly-proline II helix). NMR dynamics studies or boots-trap methods in structure calculation can help to separate conformational dynamics from the lack of experimental data. A more fundamental and not easy to overcome problem is the time and ensemble averaging during the NMR measurement.[4−9,39] NOE intensities are time averages of tens of milliseconds and ensemble averages of roughly 10^{17} molecules. Clearly, it is impossible to fully incorporate this averaging in a structure calculation. In almost all cases constraints are applied to one molecule, sometimes with some kind of picosecond time averaging. Ensemble calculations with a small number of molecules have been performed, but only in rare cases (ref 6 and references therein). The "averaging problem" is well-known in the NMR literature,[4,38,39] but there has also been a renewed discussion in recent years.[8,9] By comparing for peptides **2**, **8**, and **9** very elaborate unconstrained MD simulations with experimental NMR distance data, we have found that the bias of the NOE or ROE toward shorter distances can lead to suppression of more open and less compact structures in NMR structure calculations when the peptide is quite flexible.[40] However, the purpose of the present paper is not to validate or falsify the NMR structural ensembles for any or every peptide but to compare the performance of commonly used approaches to calculation of peptide structures taking the derived NOE distances for peptides **1**−**20** as given. The results presented above have pointed to the fact that conformational sampling is the most important aspect that needs to be considered. The objective of this work can, therefore, be stated as the following: Do the protocols investigated here explore the conformational space sufficiently. The aim is to test the performance of the computational approach, not the validity

of the individual NOE data or, even, final structures. Among our set of test molecules are some that are known, or likely, to exhibit internal motions and some that do not. Some of either kind displayed energy barriers that render the MD approach inappropriate. The lack of correlation between the presence of an energy barrier and characteristics such as ring size, additional constraints or even similarity of structures suggests that our findings are quite general. They are also in agreement with an emerging consensus that for peptides and proteins the energy landscape in conformational space is rough and rugged. Our test set of molecules comprises almost exclusively mono- or bicyclic peptides of moderate size. However, unconstrained linear peptides usually are mostly unstructured in solution, unless they experience the structuring influence of, for instance, a membrane environment (as does peptide **12**) or of a binding partner. Further, experimental NMR restraints will act to constrain the molecule when used in the structure calculation. Therefore, we believe that investigation of constrained peptides does not constitute a limitation, which is too severe. For the discussion of the performance of the various calculation protocols we will begin with those that use a given starting structure and generate the ensemble by conformational searching during a molecular dynamics simulation (protocols #2 and #3). It is clear that the energy barriers that impede proper sampling are the main problem of the conformational search. Table 1 shows that for many peptides even a simulation temperature of 1000 K was insufficient for overcoming the barriers. In our cases all calculations started from low energy structures fulfilling the NMR restraints and, thus, ensembles were found that represented subsets of the corresponding DG ensembles. However, in the real case the initial structure might be far from the correct ensemble in conformational space, and, thus, trapping of the conformational search by barriers could result in completely artificial results not corresponding to lowest energy structures. Furthermore, in additional calculations with a different force field (AMBER) we found a dependence of the results and also the existence of energy barriers on the force field used (data not shown). This indicates that for a given peptide the problem of energy barriers in MD conformational sampling might depend on the force field and MD software used. Unfortunately, peptide structure determinations based on protocols similar to our protocol 2 are seen quite often in the literature. With our demonstration of the shortcomings of these protocols we would like to discourage the use of them. In protocol 3 scaling of the nonbonded interactions partially alleviates the problems associated with protocol 2. However, in a few cases conformational sampling is still insufficient, and sometimes performance in terms of NOE violations or final energies is not satisfactory. Scaling of force field interactions is more typical for protein structure determination than for peptide studies. Although our study is limited to peptides, we generally recommend against the use of protocol 3 for the structure calculation of peptides or proteins.

Having seen that in many cases MD does not adequately explore conformational space, the question arises whether the restrictions originate from the force field or the NMR constraints that are applied. As mentioned above energy barriers seem dependent on the force field (CVFF vs AMBER). On the other hand, structure calculations without NMR restraints resulted in similar ensembles for protocols 1 and 2 for all peptides except **4 cis** (data not shown). The structures obtained without NMR data were distinctly different from those resulting from calculations with NMR restraints demonstrating that our molecules are not conformationally trivial in the sense that their conformation would already be determined by steric requirements of the amino acid residues and the intramolecular cyclization. Because NOEs contribute to the energy of the molecule they also modify the energy landscape and can create barriers. Apparently NMR constraints are more important than the force field with regard to the presence of energy barriers, although both contribute. The importance of the NOE constraints and the fact that they are very similarly implemented in the various programs for NMR structure calculation suggests that simply moving to another program or another force field might not solve the problems discussed here.

It might be argued that free MD combined with subsequent structure selection based on NMR data represents the optimum solution and in fact this method is also frequently seen in published studies. Aside from the remaining danger of incomplete sampling, as seen for peptide **4 cis** in our study, the yield in low energy structures conforming to the NMR constraints is usually reduced, so that many more structures have to be calculated to obtain a statistical ensemble.

The DG based approach is not restricted by the presence of energy barriers as independent structures are generated by a direct geometrical method.[12] This does not a priori guarantee that every conformation that is compatible with the experimental distance constraints will be found, i.e., that the conformational sampling is sufficient. But the results shown here clearly indicate that DG performs much better than high-temperature MD with regard to sampling properties. In fact, the agreement between the structural ensembles obtained by the DG/MD protocol and the random MD protocol suggests that likely no solutions to the structure determination problems were overlooked by the DG method (see below). Of course, a proper choice of parameters, or algorithms in the case of DG, is a prerequisite for every calculation strategy. We have used reasonable implementations of each method rather than ideal ones, because we intended to focus on the applicability for the nonexpert user.

As DG structures exhibit poor covalent geometry, subsequent refinement by MD simulated annealing steps is indispensable.[41−43] A change of force field (AMBER instead of CVFF) has only little consequences for the resulting structures (data not shown), because the molecular dynamics in this approach only serve to establish a correct local (covalent) geometry with corresponding relaxation of potential energies We noticed that the final energies of the DG ensembles depend somewhat on the extent of annealing that is performed (see above) and are partially higher than those obtained with the pure MD protocol (Figure 6). Although energy constitutes the sorting criterion in our analysis, it cannot be expected that values obtained at our level of sophistication (homogeneous dielectric constant, no cross

terms between potentials, etc.) will be accurate enough for a quantitative discussion. Still it might be of interest that such small peptide models seem to exhibit energy landscapes with pronounced roughness complicating the annealing procedure. Certainly an advantage of the DG/MD protocol is that the DG part that generates (and roughly optimizes) the structural ensemble requires for peptides much less computational time than the other protocols.

For proteins, on the other hand, the DG/MD approach was found to be less efficient,[41,42] so that the typical strategy for protein structure calculation is Protocol 4 that consists of iterative annealing of random starting structures in the presence of NMR restraints. The random coordinates require initially strongly reduced force field interactions that are restored to their normal value in the course of the MD steps. Protocol 4 gave quite similar results as protocol 1 (DG/MD) for our peptides. No additional conformational families were found with the random MD approach compared to the DG based calculations. The fact that the same conformational families are detected by the DG method and the randomization of coordinates suggests that for our peptides both procedures achieve sufficient sampling of the conformational space. The somewhat higher conformational variability of the random MD ensembles (Figure 3) is probably related to the higher overall energies of the final structures (Figure 6). If an ensemble cannot be adequately refined by the annealing step, that is, higher energies are found for the same overall conformations and similar degrees of NOE violation, we assume that the initially generated structures were less compatible with the force field. In a real structure determination one would need to achieve a fully relaxed conformational ensemble by extended or repeated simulated annealing. The increased computational costs compared to the other protocols lead us to consider protocol 4 as less effective. The higher efficiency of the DG based methods in this regard is intuitively understood, because DG generates independent, but not arbitrary structures, that exhibit roughly correct covalent geometry and satisfy the experimental NMR data. Random structures have to acquire these properties during the MD simulated annealing, while the DG structures are merely refined during the MD part.

Conclusions

MD simulations employing experimental NMR restraints are not well suited for conformational searching in a structure calculation protocol for peptides. We observed artificially well-defined structural ensembles for our test set of 20 peptides often accompanied by energy barriers that could not be overcome during 1000 ps simulation runs at 1000 K. The results of the pure MD method for structure calculation were found to depend markedly on the force field used. Distance geometry based protocols on the other hand explore the conformational space more thoroughly and are quite insensitive to changes in force field parameters as they utilize molecular dynamics only for refinement. We recommend the combined DG/MD approach as the most efficient method for the calculation of peptide structures.

Acknowledgment. Dr. Marion Götz is acknowledged for her help in the preparation of the manuscript. This study was supported in part by the Deutsche Forschungsgemeinschaft.

References

(1) The PDB team, The Protein Data Bank. In *Structural Bioinformatics*; Bourne, P. E., Weissig H., Eds.; John Wiley & Sons: Hoboken, NJ, 2002; pp 181−198.

(2) Wider, G. *Prog. Nucl. Magn. Reson. Spectrosc.* **1998**, *32*, 193−275.

(3) Lee, S. C.; Russell, A. F.; Laidig, W. D. *Int. J. Pept., Prot. Res.* **1990**, *35*, 367−377.

(4) Constantine, K. L.; Mueller, L.; Andersen, N. H.; Tong, H.; Wandler, C. F.; Friedrichs, M. S.; Bruccoleri, R. E. *J. Am. Chem. Soc*. **1995**, *117*, 10841−10854.

(5) Melacini, G.; Zhu, Q.; Goodman, M. *Biochemistry* **1997**, *36*, 1233−1241.

(6) Cuniasse, P.; Raynal, I.; Yiotakis, A.; Dive, V. *J. Am. Chem. Soc.* **1997**, *119*, 5239−5248.

(7) Daura, X.; Antes, I. van Gunsteren, W. F.; Thiel, W.; Mark, A. E. *Proteins* **1999**, *36*, 542−555.

(8) Bürgi, R.; Ritera, J.; van Gunsteren, W. F. *J. Biomol. NMR* **2001**, *19*, 305−320.

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

(10) Gnanakaran, S.; Nymeyer, H.; Portman, J.; Sanbonmatsu, K. Y.; Garcia, A. E. *Curr. Opin. Struct. Biol*. **2003**, *13*, 168−174.

(11) Havel, T. F.; Wüthrich, K. *J. Mol. Biol*. **1985**, *182*, 281−294.

(12) Crippen, G. M.; Havel, T. F. *Distance Geometry and Molecular Conformation*; Research Studies Press: Somerset, England, and John Wiley: New York, 1988.

(13) Berndt, K. D.; Guntert, P.; Wüthrich, K. *Proteins* **1996**, *24*, 304−313.

(14) Wells, C.; Glunt, W.; Hayden, T. L. *J. Mol. Struct*. **1994**, *308*, 263−271.

(15) Hodsdon, M. E.; Ponder, J. W.; Cistola, D. P. *J. Mol. Biol.* **1996**, *264*, 585−602.

(16) Reams, R.; Chatham, G.; Glunt, W.; McDonald, D.; Hayden, T. *Comput. Chem.* **1999**, *23*, 153−163.

(17) Vankampen, A. H. C.; Buydens, L. M. C.; Lucasius, C. B.; Blommers, M. J. J. *J. Biomol. NMR* **1996**, *7*, 214−224.

(18) Huang, E. S.; Samudrala, R.; Ponder, J. W. *J. Mol. Biol*. **1999**, *290*, 267−281.

(19) Xu, H. F.; Izrailev, S.; Agrafiotis, D. K. *J. Chem. Inf. Comput. Sci*. **2003**, *43*, 1186−1191.

(20) Havel, T. F. *Biopolymers* **1990**, *29*, 1565−1585.

(21) Edwards, J.; Chatham, G.; Glunt, W.; McDonald, D.; Wells, C.; Hayden, T. *Comput. Chem*. **1997**, *21*, 115−124.

(22) Vankampen, A. H. C.; Beckers, M. L. M.; Buydens, L. M. C. *Comput. Chem*. **1997**, *21*, 281−297.

(23) Liu, Y.; Zhao, D.; Altman, R. Jardetzky, O. *J. Biomol. NMR* **1992**, *2*, 373−388.

(24) Beckers, M. L. M.; Buydens, L. M. C.; Pikkemaat, J. A.; Altona, C. *J. Biomol. NMR* **1997**, *9*, 25−34.

(25) Karimi-Nejad, Y.; Warren, G. L.; Schipper, D.; Brünger, A. T.; Boelens, R. *Mol. Phys.* **1998**, *95*, 1099−1112.

(26) Nilges, M.; Gronenborn, A. M.; Brünger, A. T.; Clore, G. M. *Protein Eng.* **1988**, *2*, 27−38.

(27) Dauber-Osguthorpe, P.; Roberts, V. A.; Osguthorpe, D. J.; Wolff, J. Genest, M.; Hagler, A. T. *Proteins* **1988**, *4*, 31−47.

(28) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *J. Comput. Chem*. **1986**, *7*, 230−252.

(29) Renner, C.; Behrendt, R.; Spörlein, S.; Wachtveitl, J.; Moroder, L. *Biopolymers* **2000**, *54*, 489−500.

(30) Renner, C.; Cramer, J.; Behrendt, R.; Moroder, L. *Biopolymers* **2000**, *54*, 501−514.

(31) Renner, C.; Behrendt, R.; Heim, N.; Moroder, L. *Biopolymers* **2002**, *63*, 382−393.

(32) Schütt, M.; Krupka, S.; Milbradt, A. G.; Deindl, S.; Sinner, E.-K.; Oesterhelt, D.; Renner, C.; Moroder L. *Chem. Biol*. **2003**, *10*, 487−490.

(33) Spörlein, S.; Carstens, H.; Satzger, H.; Renner, C.; Behrendt, R.; Moroder, L.; Tavan, P.; Zinth, W.; Wachtveitl, J. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 7998−8002.

(34) Bredenbeck, J.; Helbing, J.; Sieg, A.; Schrader, T.; Zinth, W.; Renner, C.; Behrendt, R.; Moroder, L.; Wachtveitl, J.; Hamm, P. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 6452−5457.

(35) Bredenbeck, J.; Helbing, J.; Behrendt, R.; Renner, C.; Moroder, L.; Wachtveitl, J.; Hamm, P. *J. Phys. Chem. B* **2003**, *107*, 8654−8660.

(36) Wachtveitl, J.; Spörlein, S.; Satzger, H.; Fonrobert, B.; Renner, C.; Behrendt, R.; Oesterhelt, D.; Moroder, L.; Zinth W. *Biophys. J*. **2004**, *86*, 2350−2362.

(37) Schneider, T. R.; Brünger, A. T.; Nilges, M. *J. Mol. Biol*. **1999**, *285*, 727−740.

(38) Neuhaus, D.; Williamson, M. *The nuclear Overhauser effect in Structural and Conformational Analysis;* VCH Publisher: New York, 1989.

(39) Nanzer, A. P.; Poulsen, F.; van Gunsteren, W. F.; Torda, A. E. *Biochemistry* **1994**, *33*, 14503−14511.

(40) Carstens, H.; Renner, C.; Milbradt, A. G.; Moroder, L.; Tavan, P. *Biochemistry* **2005**, *44*, 4829−4840.

(41) Holak, T. A.; Kearsley, S. K.; Kim, Y.; Prestegard, J. H. *Biochemistry* **1988**, *27*, 6135−6142.

(42) Holak, T. A.; Nilges, M.; Oschkinat, H. *FEBS Lett*. **1989**, *242*, 218−224.

(43) Knegtel, R. M.; Boelens, R.; Ganadu, M. L.; Kaptein, R. *Eur. J. Biochem*. **1991**, *202*, 447−458.

CT050203J

JCTC Journal of Chemical Theory and Computation

A Fast QM/MM (Quantum Mechanical/Molecular Mechanical) Approach to Calculate Nuclear Magnetic Resonance Chemical Shifts for Macromolecules

Bing Wang and Kenneth M. Merz, Jr.*,†

104 Chemistry Building, The Pennsylvania State University, University Park, Pennsylvania 16802

Received August 23, 2005

Abstract: A fast approach to calculate nuclear magnetic resonance (NMR) chemical shifts within the quantum mechanical/molecular mechanical (QM/MM) framework has been developed. The QM treatment is based on our recently implemented MNDO/NMR method (Wang et al. *J. Chem. Phys.* **2004**, *120*, 11392). The effect of the QM/MM partitioning on chemical shifts has been investigated by test calculations on the water dimer and on the protein crambin. It has been shown that the quantum mechanical treatment of the hydrogen bond and nearby groups with significant magnetic susceptibilities is necessary in order to reproduce the full QM results. The method is also applied to a protein−ligand complex FKBP-GPI, and excellent agreement for proton chemical shifts of the ligand is obtained by including the side-chain atoms of the binding site residues into the QM region. The NMR chemical shift calculations using QM/MM-minimized structures still yield satisfactory results. Our results demonstrate that this QM/MM NMR method is able to treat critical regions of very large macromolecules without compromising accuracy if a relatively large QM region is used.

I. Introduction

Over the past several decades, nuclear magnetic resonance (NMR) spectroscopy has emerged as a powerful tool to study the structure and dynamics of biological systems.[1] One of its essential parameters is the NMR chemical shift that characterizes the chemical environment of individual atoms. Although it has been shown that the NMR chemical shift can provide useful information for protein secondary structures,[2] its complicated relationship with molecular structures hinders its application in protein structure determination and other NMR studies of biological systems. NMR chemical shifts are sensitive to subtle changes in electronic structure from local variations in bond lengths and bond and torsional angles, to electrostatic interactions, to hydrogen bonds, and to magnetic susceptibility effects. In principle, the quantum mechanical theory of chemical shieldings is able to capture all of these contributions and predict chemical shifts accurately. Indeed, it has been demonstrated that calculated ^{1}H and ^{13}C chemical shifts using reasonable basis sets at the Hartee−Fock and density functional theory (DFT) levels can predict experimental values for a variety of organic molecules with excellent accuracy.[3−5] However, the substantial expense of these methods hinders their application to macromolecules with thousands of atoms. As a result of the increase of NMR studies of biological systems, empirical methods[6−9] have been developed to reproduce measured chemical shifts by parametrization and have been shown to be useful in protein structure refinement.[10] While they have been reasonably successful for protein systems, these empirical approaches are not designed to study protein−ligand complexes because a variety of ligand molecular structures generally are not in the parametrization set. Recently, we have developed a fast approach, DCNMR,[11] to calculate NMR chemical shifts using the divide-and-conquer method[12−14] at the MNDO[15] level.

* Corresponding author phone: (352) 392-6973; fax: (352) 392-8722; e-mail: merz@psu.edu.

† Current address: University of Florida, Department of Chemistry, Quantum Theory Project, 2328 New Physics Building, P. O. Box 118435, Gainesville, FL 32611. E-mail: merz@qtp.ufl.edu.

10.1021/ct050212s CCC: $33.50
Published on Web 12/01/2005

By utilizing parameters specifically developed for NMR calculation,[16] excellent agreement with experimental results was obtained for ligand proton chemical shifts in the FK506 binding protein (FKBP)−GPI complex[17] and cellular retinol-binding protein.[18] Applications to decoy pose scoring[17] and NMR structure refinement[18] have been illustrated as well.

In this paper, we extend our DCNMR approach into the quantum mechanical/molecular mechanical (QM/MM) framework. The motivations for this work were as follows: (1) Although our original approach is able to compute NMR chemical shifts for macromolecules containing thousands of atoms, it is still computationally intensive to address these systems. Since the chemical shieldings are strongly dependent on local electronic environments, in many instances, it is unnecessary to treat all atoms in macromolecules by the more expensive quantum mechanical approach. (2) If the interest is localized in the binding site of the protein−ligand or protein−protein complex, it is natural to calculate NMR chemical shifts within the framework of QM/MM; the ligand and the residues inside the binding site are treated quantum mechanically, while the rest of system is computed molecular mechanically. (3) A good structure is a requirement for NMR chemical shift calculations. However, geometry optimization of an entire protein with several thousands of atoms is a substantial task even for linear-scaling semiempirical methods. If we can use QM/MM-optimized geometries for the NMR chemical shift calculation, it will significantly speed up the entire process.

A number of NMR chemical shift calculations using the QM/MM approach have been recently reported. Cui and Karplus[19] have combined Gaussian and CHARMM to compute NMR chemical shifts using the QM/MM approach, and this is the most general implementation so far. Their results demonstrated that the QM/MM approach can reach the same accuracy as a full QM calculation by using hydrogen atoms as link atoms. By capping the QM region with quantum capping potentials and representing the MM region with point charges, Moon et al.[20] proposed a simple approach to calculate NMR shielding tensors since most quantum chemistry programs can handle effective core potentials and point charges. Another implementation of NMR chemical shift calculation using plane wave basis sets with repulsive potentials in conjunction with the QM/MM strategy has appeared as well.[21] All these approaches utilized ab initio and DFT methods to treat the QM region, which might not be ideally suited for high-throughput screening studies of protein−ligand complexes, for example. The total number of atoms in the binding site (including the inhibitor and the residues around it) could be so large (potentially over 100 atoms) that the computational costs of current ab initio and DFT methods are still prohibitive for routine NMR chemical shift calculations. Moreover, it may be necessary to use a relatively large QM region to alleviate nonphysical effects arising from the use of link atom schemes at the boundary of the QM and MM regions. Here, we incorporate our fast DCNMR approach into AMBER to compute NMR chemical shifts using the QM/MM approach. This coupling makes it possible to provide insights into biomacromolecules using chemical shift information, such as building relationships between predicted chemical shifts and the protein structure, studying dynamical effects on chemical shifts, and even performing virtual high-throughput NMR-based screening on large sets of molecules.

II. Method and Implementation

Since our DCNMR approach has been published elsewhere,[11] only the essentials relating to the QM/MM implementation are outlined herein. The chemical shielding tensor σ_{ab} is the second derivative of the molecular energy with respect to the external magnetic field and the nuclear magnetic moment, which can be expressed in the Hamiltonian form as

$$\sigma_{ab} = \sum_{\mu\nu} \mathbf{P}_{\mu\nu} H^{ab}_{\mu\nu} - \sum_{\mu\nu} \mathbf{P}^{a}_{\mu\nu} H^{0b}_{\mu\nu} \tag{1}$$

where $\mathbf{P}_{\mu\nu}$ is the density matrix obtained from the self-consistent field (SCF) calculation, $\mathbf{P}^{a}_{\mu\nu}$ is the derivative of the density matrix with respect to the magnetic field. $H^{ab}_{\mu\nu}$ and $H^{0b}_{\mu\nu}$ are the magnetic integral elements in the gauge-including atomic orbital. Their expressions are

$$H^{0b}_{\mu\nu} = -\frac{1}{c}\left\langle \chi_\mu \left| \frac{[(\vec{r} - \vec{R}) \times \vec{\nabla}]_b}{|\vec{r} - \vec{R}|^3} \right| \chi_\nu \right\rangle \tag{2}$$

$$\begin{aligned} H^{ab}_{\mu\nu} = \frac{1}{2c^2}\Bigg\{ & (\vec{R}_\mu \times \vec{R}_\nu)_a \left\langle \chi_\mu \left| \frac{[(\vec{r} - \vec{R}) \times \vec{\nabla}]_b}{|\vec{r} - \vec{R}|^3} \right| \chi_\nu \right\rangle + \\ & \left\langle \chi_\mu \left| [(\vec{r} - \vec{R}_\mu) \times (\vec{R}_\nu - \vec{R}_\mu)]_a \frac{[(\vec{r} - \vec{R}) \times \vec{\nabla}]_b}{|\vec{r} - \vec{R}|^3} \right| \chi_\nu \right\rangle + \\ & \left\langle \chi_\mu \left| \frac{\delta_{ab}(\vec{r} - \vec{R}_\nu)(\vec{r} - \vec{R}) - (\vec{r} - \vec{R}_\nu)_b(\vec{r} - \vec{R})_a}{|\vec{r} - \vec{R}|^3} \right| \chi_\nu \right\rangle \Bigg\} \end{aligned} \tag{3}$$

The detail implementation of these integrals was described in our original paper.[11] The perturbed density matrix $\mathbf{P}^{a}_{\mu\nu}$ can be obtained by solving the coupled-perturbed Hartree−Fock (CPHF) equations. However, current implementations of this procedure are time-consuming, although some efforts to reduce this cost have appeared.[22] To take advantage of the linear-scaling divide-and-conquer method,[12−14] we adopted an alternative approach to calculate the perturbed density matrix, which was based on finite perturbation theory. The magnetic-field-dependent Fock matrix was first built and diagonalized; then, the perturbed density matrix could be approximated by

$$\mathbf{P}^{a}_{\mu\nu} \approx \frac{iP^{i}_{\mu\nu}}{B_a} \tag{4}$$

where $P^{i}_{\mu\nu}$ is the imaginary part of the density matrix. Consequently, this approach enables us to calculate the density matrix and the perturbed density matrix simultaneously using the divide-and-conquer method.

The combination of quantum mechanics and molecular mechanics is a natural approach for the study of enzyme reactions and protein−ligand interactions.[23] The active site or binding site is treated by ab initio, density functional theory, or semiempirical potentials, whereas the rest of the

system is modeled via force fields. The corresponding Hamiltonian can be divided according to

$$\hat{H}_{\text{total}} = \hat{H}_{\text{QM}} + \hat{H}_{\text{MM}} + \hat{H}_{\text{QM/MM}} \quad (5)$$

In our approach, $\hat{H}_{\text{QM}}$ can be MNDO,[15] AM1,[24] and PM3[25] semiempirical methods implemented in the program DIVCON; $\hat{H}_{\text{MM}}$ is determined by an AMBER force field. $\hat{H}_{\text{QM/MM}}$ describes the interaction between the QM and MM atoms. In general, it is written as

$$\hat{H}_{\text{QM/MM}} = -\sum_{iM}\frac{q_M}{r_{iM}} + \sum_{\alpha M}\frac{Z_\alpha q_M}{R_{\alpha M}} + \sum_{\alpha M}\left\{\frac{A_{\alpha M}}{R_{\alpha M}^{12}} - \frac{B_{\alpha M}}{R_{\alpha M}^{6}}\right\} \quad (6)$$

where the subscripts i and α refer to the QM electrons and nuclei, respectively, and M to the MM atoms. q_M is the MM partial charge. The first two terms are electrostatic terms through which the MM atoms interact with the QM electrons and nuclei, respectively. The last term is the van der Waals term.

As pointed out by Cui and Karplus,[19] the MM atoms make contributions to the chemical shielding tensors through their perturbations on both the density matrix and the perturbed density matrix. In our DCNMR approach, both the density matrix and the perturbed density matrix are obtained simultaneously by the diagonalization of the complex Fock matrix without utilizing the more expensive CPHF equations. Therefore, it is straightforward to include the MM contributions to the chemical shielding tensors by simply adding the first term in eq 6 to the Fock matrix. The method described above has been implemented into a development version of DIVCON and AMBER8.[26]

III. Results and Discussion

In this section, we apply the QM/MM DCNMR method to a number of model systems including the water dimer, crambin, and the FKBP−GPI complex. Our focus is on the validation of the QM/MM results with respect to the full QM calculations. Future work will focus on the application of this approach to biological problems.

III.1. Water Dimer. The water dimer is a simple hydrogen-bonded system for which chemical shielding changes associated with the variation of geometric parameters have been studied by full ab initio methods[27,28] and the QM/MM method.[19] To test our approach on the water dimer, a set of structures was generated by varying the distance between the two oxygen atoms. These structures were then optimized at the B3LYP/6-311++G** level. For each structure, the NMR chemical shifts of the hydrogen and oxygen atoms relative to the isolated water molecule were calculated as

$$\Delta\delta = \sigma_{\text{monomer}} - \sigma_{\text{dimer}} \quad (7)$$

In this way, the intrinsic errors associated with the methods will be largely canceled. In the QM/MM calculations, one water was treated by our MNDO/NMR approach and the other by the TIP3P model.[29] Comparisons of the chemical shifts between full QM and QM/MM methods are shown in Figure 1 for the donor, acceptor hydrogen atoms, and the oxygen atom. Note that the energy minimum of the water dimer at the B3LYP/6-311++G** level occurs at 2.90 Å. The NMR chemical shifts for the donor hydrogen decrease monotonically as the internuclear distance increases, which is consistent with previous studies.[19,27,28] However, the QM/MM curve decays much more slowly than the full QM one, and the chemical shift differences between the two curves are significant along the entire profile (as large as 0.7 ppm at the minimum energy distance). This discrepancy is due to the absence of the Pauli repulsion contribution in the QM/MM model. Similar trends are also shown for the acceptor hydrogen and the donor oxygen atom. This strongly suggests that hydrogen-bonded interactions have to be included as part of the QM region, which further confirms Cui and Karplus's conclusion.[19]

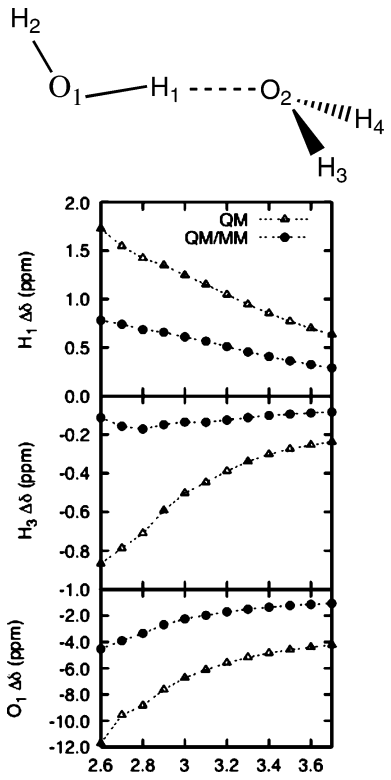


Figure 1. Comparison of the full QM and QM/MM chemical shifts for water dimers as a function of the hydrogen bonding O1···O2 distance (Å).

III.2. Crambin. Our second test system was crambin, a small hydrophobic protein with 46 residues. Both high-resolution X-ray and NMR structures are available for this protein. A small (cut 1) and a large (cut 2) QM/MM partitioning were tested for Ala9 and Pro5. As shown in Figure 2, cut 1 for Ala9 puts only the side-chain atoms in the QM region, and a link atom is introduced between the Cα and Cβ bond, while in cut 2, the QM region is extended to the neighboring peptide bonds. Figure 3 demonstrates the partition scheme for Pro5. In cut 1, the QM region is Pro5 with the neighboring peptide bonds; cut 2 is cut 1 plus a nearby residue Tyr44. The MM part is treated with the AMBER parm94 force field.[30] The entire geometry was optimized at the AM1 level before the NMR shielding calculation. Full QM NMR calculations have also been carried out as reference values using our DCNMR approach.

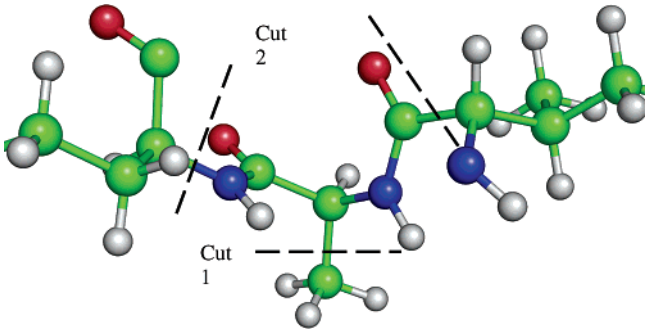


Figure 2. QM/MM partition scheme for Ala9 in crambin.

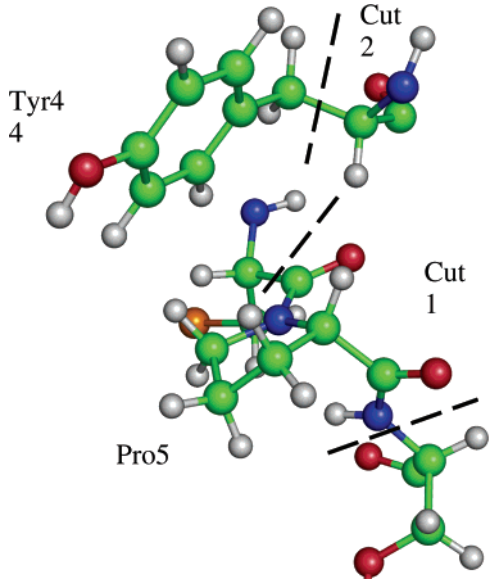


Figure 3. QM/MM partition scheme for Pro5 in crambin.

Table 1. Computed NMR Chemical Shift Errors (in ppm) at Different QM/MM Partition Schemes for Crambin Ala9 and Pro5

residue	atom	cut 1[a]	cut 2
Ala9	CB	−22.25	0.791
	HB	−1.459	0.145
Pro5	CB	0.200	−0.373
	HB2	0.613	−0.238
	HB3	0.416	0.058
	CG	0.331	−0.270
	HG2	0.535	−0.240
	HG3	0.337	0.053
	CD	−0.251	−0.623
	HD2	0.146	−0.386
	HD3	0.233	0.046

[a] Note that cut 1 and cut 2 are different for Ala9 and Pro5 (see text and Figures 2 and 3).

The differences between the QM/MM and full QM chemical shifts are summarized in Table 1. Obviously, the results for Ala9 using the cut 2 scheme are much better than that from cut 1; the error for the Cβ chemical shift is reduced from 22 to 0.7 ppm, and the errors for proton chemical shifts are also dramatically decreased from 1.5 to 0.15 ppm. This suggests that the peptide groups have significant magnetic susceptibilities and their effect on the chemical shieldings on the side-chain atoms is not negligible. Excellent agreement in cut 2 also indicates that these interactions are generally short-ranged ($1/r^3$ according to the McConnell equation[31]). Therefore, we conclude that it is realistic to treat remote peptide bonds with MM. For Pro5, the inclusion of the peptide bond atoms was able to give comparable results for the carbon atoms but was not good enough to reproduce the full QM results for hydrogen atoms. The chemical shift errors for HB and HG are 0.6 and 0.4 ppm, respectively. This problem was alleviated through the addition of the proximal residue Tyr44 (cut 2 model) to the QM NMR calculation. This reduced the calculated errors to 0.05 and 0.03 ppm, respectively. These calculations have demonstrated that it is necessary to put nearby polar and aromatic rings into the QM region to obtain satisfactory results from the QM/MM NMR method.

III.3. FKBP-GPI Complex. NMR spectroscopy has become a useful tool to study protein−ligand interactions, an essential step for structure-based drug design. This is because protein−ligand complex structures can be solved by NMR, and many NMR-based screening techniques have been developed for lead discovery and optimization such as structure−activity relationships by NMR.[32] The theoretically calculated chemical shieldings for a protein−ligand complex can aid in determining NMR structures and improve the accuracy of NMR screening techniques. We have carried out DCNMR calculations on the entire FKBP−GPI complex and obtained an excellent correlation between computed and experimental proton chemical shifts of the ligand.[17] Figure 4 illustrates the GPI ligand and the binding site residues. To investigate the effect of QM/MM partitioning on chemical shift calculation, we tested three different partition schemes: The QM region for cut 1 was just the ligand itself, and the entire protein was treated using MM. In cut 2, the QM region extends to the side-chain atoms of all residues inside the binding pocket (including Tyr26, Phe36, Asp37, Phe46, Phe48, Gln53, Val55, Ile56, Trp59, Tyr82, His87, Ile90, Ile91, Leu97, and Phe99), which results in 275 QM atoms (excluding link atoms). Cut 3 was based on cut 2 and adds all backbone atoms for the binding site residues. The differences in proton chemical shifts between the QM/MM and full QM calculations are shown in Figure 5. It is not surprising that the results from cut 1 have relatively large deviations (RMSD: 1.12 ppm; RMSD = root-mean-square deviation) because it only includes ligand atoms in the QM region. Most of the errors arise from protons on the pyrrolidine ring, which is situated in the hydrophobic pocket formed by aromatic residues Tyr26, Phe46, Trp59, and Phe99. It demonstrates that the treatment of these residues as MM point charges cannot simulate realistic ring current effects. The inclusion of these aromatic residues into the QM region in cut 2 dramatically reduces the RMSD to 0.18 ppm. However, when we added backbone atoms into the QM region in cut 3, the agreement with the full QM results was slightly worse, with a RMSD of 0.28 ppm, which might be due to an imbalanced description at the QM and MM boundary.

The previous sets of NMR calculation were based on the AM1 optimized structure of the entire complex. Since the geometry optimization of the whole protein is a time-

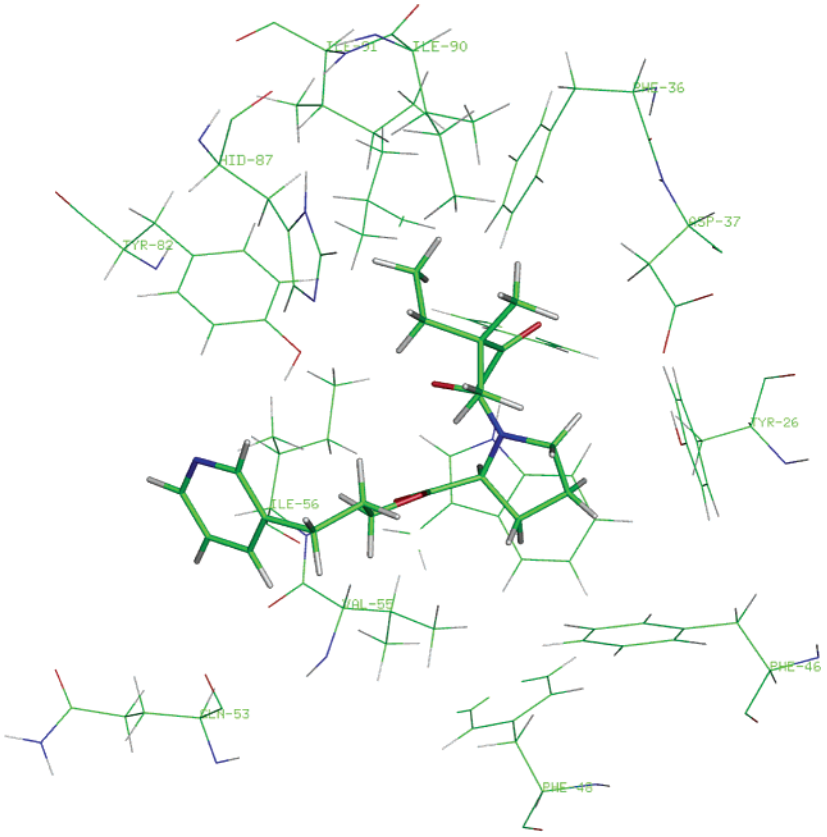


Figure 4. Binding site of FKBP−GPI complex.

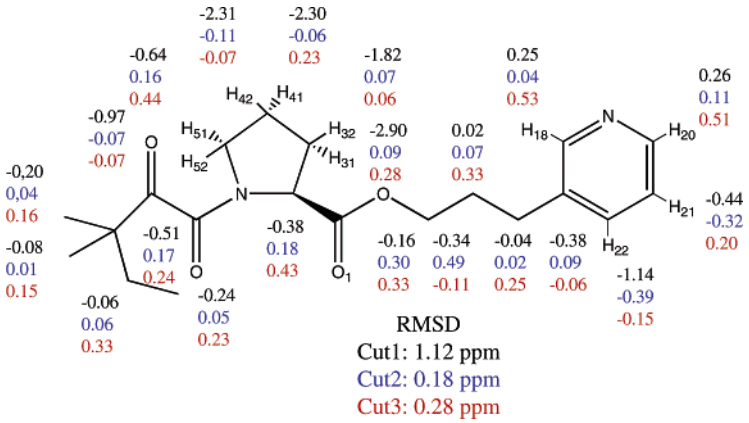


Figure 5. Chemical structure of the GPI molecule and the proton chemical shift errors for three different QM/MM partition schemes.

consuming step, another question we are trying to answer is whether we can use the less-expensive QM/MM-minimized structures for NMR calculation without compromising accuracy. Therefore, we carried out a 2000 step QM(AM1)/MM minimization of the FKBP−GPI complex based on the cut 2 scheme, followed by NMR calculation. The time saving is obvious since the total time of the QM/MM minimization was only about 2 days, while it took 8 days to optimize the entire complex by AM1 on our local opteron cluster, even with the divide-and-conquer approach. The all-atom RMSD of the binding site residues between the final geometry and full AM1 optimized geometry is only 0.7 Å. The RMSD of the calculated proton chemical shifts between the cut 2 scheme and the full QM result was 0.37 ppm. This agreement indicates that the QM/MM-minimized structures are able to produce chemical shift results that are comparable to those from fully QM-optimized structures.

IV. Conclusions

We have coupled our recently developed DCNMR approach with AMBER to calculate NMR chemical shifts within the QM/MM framework. Because our approach is able to obtain both the density matrix and the perturbed density matrix in the SCF step, it is straightforward to include the MM contributions to our NMR chemical shift calculations. Application to the water dimer indicates that it is necessary to treat the whole hydrogen bond quantum mechanically. To

investigate the QM/MM partition effect on chemical shifts of a specific residue in a protein, we have carried out calculations on two residues of crambin using different QM/MM partitioning schemes. Good agreement with full QM results was obtained when the QM region included the nearby groups with significant magnetic susceptibilities such as peptide bonds and aromatic rings. Finally, the method was applied to a protein−ligand complex (FKBP−GPI), and the proton chemical shifts of the ligand were well reproduced when the side-chain atoms of the binding site residues were included in the QM region. We have also shown that the QM/MM-optimized structure is good enough to yield satisfactory results for NMR chemical shift calculations at a fraction of the cost of full QM geometry optimization. All these test calculations have demonstrated that our QM/MM NMR method with an appropriate QM/MM partition is able to obtain good agreement with full QM results at a much lower cost and, thus, paves a way to compute NMR chemical shifts for much larger macromolecules.

Acknowledgment. We thank the NSF (MCB-0211639) and the NIH (GM 44974) for support.

References

(1) Wuthrich, K. *NMR of Proteins and Nucleic Acids*; Wiley: New York, 1986.

(2) Spera, S.; Bax, A. Empirical Correlation between Protein Backbone Conformation and C−Alpha and C−Beta C-13 Nuclear-Magnetic-Resonance Chemical-Shifts. *J. Am. Chem. Soc.* **1991**, *113* (14), 5490−5492.

(3) Rablen, P. R.; Pearlman, S. A.; Finkbiner, J. A comparison of density functional methods for the estimation of proton chemical shifts with chemical accuracy. *J. Phys. Chem. A* **1999**, *103* (36), 7357−7363.

(4) Wang, B.; Fleischer, U.; Hinton, J. F.; Pulay, P. Accurate prediction of proton chemical shifts. I. Substituted aromatic hydrocarbons. *J. Comput. Chem.* **2001**, *22* (16), 1887−1895.

(5) Wang, B.; Hinton, J. F.; Pulay, P. Accurate prediction of proton chemical shifts. II. Peptide analogues. *J. Comput. Chem.* **2002**, *23* (4), 492−497.

(6) Osapay, K.; Case, D. A. A New Analysis of Proton Chemical-Shifts in Proteins. *J. Am. Chem. Soc.* **1991**, *113* (25), 9436−9444.

(7) Williamson, M. P.; Kikuchi, J.; Asakura, T. Application of 1H NMR chemical shifts to measure the quality of protein structures. *J. Mol. Biol.* **1995**, *247* (4), 541−546.

(8) Wishart, D. S.; Watson, M. S.; Boyko, R. F.; Sykes, B. D. Automated 1H and 13C chemical shift prediction using the BioMagResBank. *J. Biomol. NMR* **1997**, *10* (4), 329−336.

(9) Xu, X. P.; Case, D. A. Automated prediction of 15N, 13Calpha, 13Cbeta and 13C′ chemical shifts in proteins using a density functional database. *J. Biomol. NMR* **2001**, *21* (4), 321−333.

(10) Osapay, K.; Theriault, Y.; Wright, P. E.; Case, D. A. Solution structure of carbonmonoxy myoglobin determined from nuclear magnetic resonance distance and chemical shift constraints. *J. Mol. Biol.* **1994**, *244* (2), 183−197.

(11) Wang, B.; Brothers, E. N.; Van Der Vaart, A.; Merz, K. M. Fast semiempirical calculations for nuclear magnetic resonance chemical shifts: A divide-and-conquer approach. *J. Chem. Phys.* **2004**, *120* (24), 11392−11400.

(12) Yang, W. T.; Lee, T. S. A Density-Matrix Divide-and-Conquer Approach for Electronic-Structure Calculations of Large Molecules. *J. Chem. Phys.* **1995**, *103* (13), 5674−5678.

(13) Dixon, S. L.; Merz, K. M. Semiempirical molecular orbital calculations with linear system size scaling. *J. Chem. Phys.* **1996**, *104* (17), 6643−6649.

(14) Dixon, S. L.; Merz, K. M. Fast, accurate semiempirical molecular orbital calculations for macromolecules. *J. Chem. Phys.* **1997**, *107* (3), 879−893.

(15) Dewar, M. J. S.; Thiel, W. Ground State of Molecules. 38. The MNDO Method. Approximations and Parameters. *J. Am. Chem. Soc.* **1977**, *99* (15), 4899−4907.

(16) Patchkovskii, S.; Thiel, W. NMR chemical shifts in MNDO approximation: Parameters and results for H, C, N, and O. *J. Comput. Chem.* **1999**, *20* (12), 1220−1245.

(17) Wang, B.; Raha, K.; Merz, K. M., Jr. Pose scoring by NMR. *J. Am. Chem. Soc.* **2004**, *126* (37), 11430−11431.

(18) Wang, B.; Merz, K. M., Jr. Validation of the binding site structure of the cellular retinol-binding protein (CRBP) by ligand NMR chemical shift perturbations. *J. Am. Chem. Soc.* **2005**, *127* (15), 5310−5311.

(19) Cui, Q.; Karplus, M. Molecular properties from combined QM/MM methods. 2. Chemical shifts in large molecules. *J. Phys. Chem. B* **2000**, *104* (15), 3721−3743.

(20) Moon, S.; Christiansen, P. A.; DiLabio, G. A. Quantum capping potentials with point charges: A simple QM/MM approach for the calculation of large-molecule NMR shielding tensors. *J. Chem. Phys.* **2004**, *120* (19), 9080−9086.

(21) Sebastiani, D.; Rothlisberger, U. Nuclear magnetic resonance chemical shifts from hybrid DFT QM/MM calculations. *J. Phys. Chem. B* **2004**, *108* (9), 2807−2815.

(22) Weber, V.; Niklasson, A. M. N.; Challacombe, M. *Phys. Rev. Lett.* **2004**, *92* (19), 193002.

(23) Warshel, A.; Levitt, M. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* **1976**, *103*, 227−249.

(24) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. AM1: A New General Purpose Quantum Mechanical Molecular Model. *J. Am. Chem. Soc.* **1985**, *107* (13), 3902−3909.

(25) Stewart, J. J. P. Optimization of Parameters for Semiempirical Methods I. Method. *J. Comput. Chem.* **1989**, *10* (2), 209−220.

(26) Case, D. A.; Darden, T. A.; Cheatham, I., T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Wang, B.; Pearlman, D. A.; Crowley, M.; Brozell, S.; Tsui, V.; Gohlke, H.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Schafmeister, C.; Caldwell, J. W.; Ross, W. S.; Kollman, P. A. *AMBER*, 8.0; 2004.

(27) Dithfield, R. Theoretical studies of magnetic shieldingin H_2O and $(H_2O)_2$. *J. Chem. Phys.* **1976**, *65*, 3123−3133.

(28) Chesnut, D. B.; Rusiloski, B. E. Partial energy and chemical shielding surfaces in the water (H2O)2 and hydrogen fluoride (HF)2 van der Waals complexes. *J. Phys. Chem.* **1993**, *97*, 2839−2845.

(29) 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.

(30) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. C.; Spellmeyer, T.; Fox, J. W.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Protein, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.

(31) McConnell, H. M. Theory of Nuclear Magnetic Shielding in Molecules. I. Long-Range Dipolar Shielding of Protons. *J. Chem. Phys.* **1957**, *27*, 226−229.

(32) Shuker, S. B.; Hajduk, P. J.; Meadows, R. P.; Fesik, S. W. Discovering high-affinity ligands for proteins: SAR by NMR. *Science* **1996**, *274*, 1531−1534.

CT050212S

JCTC Journal of Chemical Theory and Computation

Improving the QM/MM Description of Chemical Processes: A Dual Level Strategy To Explore the Potential Energy Surface in Very Large Systems. [*J. Chem. Theory Comput. 1*, 1008−1016 (2005)]. By Sergio Martí, Vicente Moliner, and Iñaki Tuñón. Departament de Ciències Experimentals, Universitat Jaume I, Box 224, 12080 Castellón, Spain and Departament de Química Física/IcMol, Universidad de Valencia, 46100 Burjasot, Valencia, Spain

Pages 1013 and 1014. In the last example discussed in the paper we used a combination of B3LYP/6-31+G* and PM3 methods combined with molecular mechanics (MM) named as B3LYP:PM3/MM. At the end of the section (p 1014) we erroneously named it twice as B3LYP:AM1/MM, instead of B3LYP:PM3/MM. This typographical correction has no bearing with any of our results or conclusions.

CT058001A

10.1021/ct058001a
Published on Web 12/02/2005