Perturbation Approach to Combined QM/MM Simulation of Solute−Solvent Interactions in Solution

Elena Cubero,[†] F. Javier Luque,*[,‡] Modesto Orozco,*[,†] and Jiali Gao*[,§]

Departament de Bioquímica. Facultat de Química, Universitat de Barcelona, Martí i Franquès 1, Barcelona 08028, Spain, Departament de Fisicoquímica, Facultat de Farmàcia, Universitat de Barcelona, Avgda. Diagonal s/n, Barcelona 08028, Spain, and Department of Chemistry and Minnesota Supercomputer Institute, University of Minnesota, 207 Pleasant Street, SE, Minneapolis, Minnesota 55455

Received: August 29, 2002; In Final Form: December 6, 2002

A perturbation approach based on the generalized molecular interaction potential with polarization (GMIPp) has been developed for combined quantum mechanical and molecular mechanical (QM/MM) simulations of solutions. A unique feature of the method is to avoid repeated self-consistent field calculations for each new solvent configuration during the fluid simulation. The results show that the GMIPp potential coupled to Monte Carlo simulations yields similar energetic and structural results for a series of 14 organic compounds in aqueous solution in comparison with the results obtained from full ab initio QM/MM simulations at the HF/6-31G(d) level. Importantly, the solute electronic polarization energy is reasonably estimated by the GMIPp method. The implementation of the GMIPp within the Monte Carlo simulation framework appears to be a promising computational strategy for carrying out QM/MM calculations for large molecules, which otherwise would be difficult by using standard QM/MM techniques.

I. Introduction

Combined quantum mechanical and molecular mechanical (QM/MM) potentials coupled with Monte Carlo or molecular dynamics simulations offer the opportunity to accurately study chemical processes in solutions and in enzymes.[1−6] A major attractive feature of this approach is that only a small portion of a large system is treated explicitly by quantum mechanics, while the rest of the system is approximated by molecular mechanics force field. Consequently, quantum mechanical calculations can be carried out on large molecules, enabling a wide range of problems of chemical and biological interest to be studied.[2,7,8] However, it is still time-consuming to use ab initio molecular orbital or density functional theory (DFT) in QM/MM simulations because a very large number of electronic structure calculations (millions in Monte Carlo algorithms) are needed to achieve adequate statistical sampling.[9−11] Therefore, it is useful to explore alternative methods that can reduce the computational cost in condensed phase QM/MM simulations.[12−16]

In this article, we present a comparative study of the solvation of a series of organic solutes in water, employing a combined ab initio QM/MM method,[17] and the generalized molecular interaction potential with polarization correction (GMIPp),[18−20] both at the HF/6-31G(d) level. In the latter calculation, a second-order perturbation approach is used to compute the polarization energy between QM and MM regions.[20] As a result, self-consistent field (SCF) iterations are avoided in QM/MM energy calculations when positions of solvent molecules are changed. The GMIPp-QM/MM simulation method is computationally efficient, yet it retains similar accuracy in comparison with standard QM/MM methods, and proves to be very well suited for Monte Carlo QM/MM simulations.

II. Method

In this section, we first summarize the energy components in combined QM/MM calculations. We then present the theoretical framework for the generalized molecular interaction potential with polarization energy, which is compared with explicit, full QM/MM simulations. In this article, we do not address the polarization of the MM region in the presence of a QM region and the reset of the MM electric field.[21,22]

II.A. Combined QM/MM Energy Components. The effective Hamiltonian for a QM/MM system is partitioned into a QM and an MM region, which consists of three major components:[1,3] (1) the electronic Hamiltonian of the isolated "QM" molecule in the gas phase, $\hat{H}_{qm}$, (2) the potential energy due to interactions among MM molecules, $\hat{H}_{mm}$, and (3) the coupling term between the QM and MM regions, $\hat{H}_{qm/mm}$:

$$\hat{H}_{eff} = \hat{H}_{qm} + \hat{H}_{qm/mm} + \hat{H}_{mm} \quad (1)$$

The first two terms in eq 1 contain electronic degrees of freedom of the QM region and, thus, are explicitly included in the self-consistent field (SCF) calculations, which are performed to determine the molecular wave function of the QM molecule. It should be noted that to account for long-range dispersion interactions and short-range exchange repulsion between the QM and MM molecules, a simple Lennard-Jones term is typically introduced. Consequently, $\hat{H}_{qm/mm}$ adopts the form given in eq 2

$$\hat{H}_{qm/mm} = \hat{H}^{elec}_{qm/mm} + \hat{H}^{vdW}_{qm/mm} \quad (2)$$

where $\hat{H}^{elec}_{qm/mm}$ is the electrostatic interaction Hamiltonian between MM partial charges and electrons and nuclei in the QM

* To whom correspondence should be addressed.
[†] Departament de Bioquímica. Facultat de Química, Universitat de Barcelona.
[‡] Departament de Fisicoquímica, Facultat de Farmàcia, Universitat de Barcelona.
[§] University of Minnesota.

10.1021/jp026874k CCC: $25.00
Published on Web 01/23/2003

region and $\hat{H}^{vdW}_{qm/mm}$ is the nonelectrostatic component (van der Waals interaction) between QM and MM atoms.[1,3]

The total potential energy of the QM/MM system is given by eq 3

$$E_{tot} = E_X + E^{elec}_{Xs} + E^{vdW}_{Xs} + E_{mm} \quad (3)$$

where E_X is the energy of the QM molecule in the gas phase at the geometry as it is in solution, E_{mm} is the interaction energy of the solvent molecules, and E^{elec}_{Xs} and E^{vdW}_{Xs} are electrostatic and van der Waals interaction energies between the solute and solvent (or QM and MM) molecules. Here, we have used the subscript X to specify the QM solute molecule and s to denote solvent.

We define the solute−solvent, or QM/MM, *electrostatic* interaction energy as follows:

$$\Delta E^{elec}_{Xs} = \langle\Psi|\hat{H}_{qm} + \hat{H}^{elec}_{qm/mm}|\Psi\rangle - \langle\Psi^0|\hat{H}^{qm}|\Psi^0\rangle \quad (4)$$

where the wave functions Ψ^0 and Ψ are, respectively, defined by $\hat{H}_{qm}|\Psi^0\rangle = E^O_X|\Psi^0\rangle$ and $\hat{H}_{eff}|\Psi\rangle = E_{tot}|\Psi\rangle$, corresponding to the molecular wave function in the gas phase and in solution. It is clear from eq 4 that ΔE^{elec}_{Xs} is the energy difference for the QM solute in solution and in the gas phase. With this definition, the total potential energy of the system (eq 3) can also be written as

$$E_{tot} = E^O_X + \Delta E^{elec}_{Xs} + E^{vdW}_{Xs} + E_{mm} \quad (5)$$

Equation 5 is useful because it separates the electrostatic energy associated with the solute into a term corresponding to the solute energy in the gas phase, $E^O_X = \langle\Psi^0|\hat{H}_{qm}|\Psi^0\rangle$, and an electrostatic interaction term due to interactions with the solvent, ΔE^{elec}_{Xs}.

Alternatively, eq 4 can also be expressed in terms of perturbation theory by treating the solute−solvent interaction as a perturbation to the solute wave function:[23]

$$\Delta E^{elec}_{Xs} = E^{(1)} + E^{(2)} + E^{(3)} + ... \quad (6)$$

The first-order perturbation term $E^{(1)}$ in eq 6 is the interaction energy between the "unpolarized" solute and MM solvent, called the vertical interaction energy[22,23]

$$\Delta E^{ver}_{Xs} = E^{(1)} = \langle\Psi^0|\hat{H}^{elec}_{qm/mm}|\Psi^0\rangle \quad (7)$$

The ΔE^{ver}_{Xs} term corresponds to the energy change to transfer the solute from the gas phase to solution, keeping the solute charge distribution fixed at its gas-phase value. The vertical interaction energy can be evaluated using the gas-phase density matrix and QM/MM one-electron integrals.

All higher order perturbation terms in eq 6 include modifications to the gas-phase solute wave function, which by definition contribute to the electronic polarization energy and can be formally defined as[22,23]

$$\Delta E^{pol}_{Xs} = E^{(2)} + E^{(3)} + ... \quad (8)$$

or, equivalently

$$\Delta E^{pol}_{Xs} = \langle\Psi|\hat{H}_{qm} + \hat{H}^{elec}_{qm/mm}|\Psi\rangle - \langle\Psi^0|\hat{H}_{qm} + \hat{H}^{elec}_{qm/mm}|\Psi^0\rangle \quad (9)$$

In this decomposition analysis, the QM/MM electrostatic interaction is the sum of the vertical interaction term and the polarization energy

$$\Delta E^{elec}_{Xs} = \Delta E^{ver}_{Xs} + \Delta E^{pol}_{Xs} \quad (10)$$

It is important to note that, in QM/MM calculations based on the HF framework, the solute wave functions (in gas phase and solution) are determined by the full SCF method for each configuration sampled during molecular dynamics or Monte Carlo simulations. Therefore, it is not necessary to use the perturbation formula of eq 8 to determine the ΔE^{pol}_{Xs} term. Instead, eq 9 is used. However, fully converged SCF calculations are time-consuming and are often the computational bottleneck in combined QM/MM methods, particularly if ab initio or DFT methods are used to describe the QM part of the system.[17] Consequently, it is highly desirable to evaluate methods that can speed up combined QM/MM calculations.[12,13]

II.B. Generalized Molecular Interaction Potential with Polarization Energy. From eq 9, it is clear that major computational efforts in combined QM/MM methods are spent to determine the polarization energy of the QM system that is induced by the instantaneous configurations of the MM electric field. This is because the vertical interaction energy (eq 7) depends only on the charge density of the solute molecule in the gas phase. Thus, SCF calculations only need to be performed once. Specifically, eq 7 can be explicitly expressed as follows:

$$E^{ver}_{qm/mm} = \sum_{m=1}^{M} q_m V(\mathbf{R}_m) \quad (11)$$

where M is the total number of MM atoms. In eq 11, $V(\mathbf{R}_m)$ is the molecular electrostatic potential (MEP) from the QM molecule at position $\mathbf{R}_m$, where the MM charge q_m, $m = 1, ..., M$, is located. The MEP can be determined conveniently from HF−SCF calculations:[18,19,24]

$$V(R_m) = \sum_{a=1}^{A} \frac{Z_a}{R_{ma}} - \sum_{\mu,\nu} P^o_{\mu\nu} I_{\mu\nu}(\mathbf{R}_m) \quad (12)$$

where A is the total number of QM atoms, Z_a is the nuclear charge of atom a, R_{ma} is the distance between QM atom a and MM atom m, and $P^o_{\mu\nu}$ and $I_{\mu\nu}$ ($\mathbf{R}_m$) are elements of the usual density and one-electron integral matrix for a unit charge located at $\mathbf{R}_m$. The superscript "o" indicates that the density matrix is computed from the gas-phase wave function of the solute molecule, and the one-electron integral is defined in eq 13, where ϕ is an atomic orbital basis function:

$$I_{\mu\nu}(\mathbf{R}_m) = \left\langle \phi_\mu \left| \frac{1}{|\mathbf{R}_m - \mathbf{r}_1|} \right| \phi_\nu \right\rangle \quad (13)$$

If the computation of the electronic polarization energy given in eq 8 is truncated at the second-order perturbation level, it can be approximated as follows:[20]

$$\Delta E^{pol}_{Xs} \approx E^{(2)} = \langle\Psi^0|\hat{H}^{elec}_{qm/mm}|\Psi^{(1)}\rangle = \sum_{j}^{vir}\sum_{i}^{occ} \frac{1}{\epsilon_i - \epsilon_j} \left(\sum_{\mu,\nu} c_{\mu i} c_{\nu j} \sum_{m=1}^{M} I_{\mu\nu}(\mathrm{R}_m)\right)^2 \quad (14)$$

where $\Psi^{(1)}$ is the first-order perturbed wave function, the indices i and j run through occupied and virtual molecular orbitals, respectively, ϵ_i and ϵ_j are the corresponding orbital energies, and $c_{\mu i}$ are orbital coefficients. An advantage of using eq 14 to

evaluate the solute electrostatic polarization energy is that the solute wave function in solution is no longer needed in computing the total potential energy of the system (eqs 1 and 3). Here, all that is required is the molecular orbital coefficients for the solute molecule in the gas phase, which can be obtained only once at the start of a Monte Carlo simulation. Of course, the one-electron integrals still must be computed when solute and solvent positions are changed during the simulation.

The use of eqs 11 and 14 along with the QM/MM van der Waals term to approximate the total solute−solvent interaction energy has been referred to as the generalized molecular interaction potential with polarization correction (GMIPp).[20] Through a series of studies of hydrogen bonding interactions and comparison with polarization energies computed using SCF HF/6-31G(d),[20] the GMIPp was shown to yield excellent results and the GMIPp polarization energies deviate only 3% from full SCF results with an RMS error of only 0.7 kcal/mol for a range of polarization energies spanning over 30 kcal/mol. The GMIPp method has also been applied to the study of biopolymer interactions including aromatic stacking and cation−π interactions.[25,26] These studies support the reliability of the GMIPp, suggesting that GMIPp may be used as an alternative approach to evaluate solute polarization energies rather than using fully converged QM/MM methods. This expectation is demonstrated to be practical in this study through Monte Carlo simulations of a series of organic solutes in water.

II.C. Computational Details. The GMIPp method has been implemented into a computer program developed in our laboratory for combined QM/MM Monte Carlo simulations.[27,28] In the present study, electronic structure calculations are performed using a locally modified version of the GAMESS program,[29] interfaced with the Monte Carlo program.[28] Although the use of eq 14 only requires the gas-phase density matrix, which avoids full SCF iterations for each Monte Carlo move, it can still be inefficient if all occupied and virtual orbitals are looped over in polarization energy calculations.[20] Consequently, we have used an energy criterion to truncate the number of terms in eq 14. In the present study, we have used a value of $\Delta\epsilon = 3$ hartree, which has been found to yield polarization energies essentially (with an error of less than 1%) identical to the values calculated when all terms are included. Note that, in the present perturbation approach, the solute wave function (in gas phase) is still recomputed for every solute movement or a change in the volume of the simulated system, although this could be avoided if a rotation transformation is made to the orbital coefficients and electronic integrals if solute moves only involve translation and rotation.

Statistical mechanical Monte Carlo simulations have been carried out for systems consisting of one solute embedded in a cubic box of 260 water molecules (roughly $20 \times 20 \times 20$ Å^3).[28] We have selected a total of fourteen simple organic compounds in the investigation: water (H_2O), methanol (CH_3OH), formaldehyde (H_2CO), acetone (CH_3COCH_3), dimethyl ether (CH_3OCH_3), methylamine (CH_3NH_2), acetonitrile (CH_3CN), acetic acid (CH_3CO_2H), formamide ($HCONH_2$), *N*-methyl acetamide ($CH_3CONHCH_3$), pyridine (C_5H_5N), imidazolium ion ($C_3N_2H_5^+$), methoxide ion (CH_3O^-), and acetate ion ($CH_3CO_2^-$). In these calculations, each solute molecule is treated at the Hartree−Fock level using the 6-31G(d) basis set along with a set of standard van der Waals parameters for ab initio QM/MM potentials.[17] To test the importance of diffuse functions for the two anions and the stability of the present GMIPp method when diffuse functions are used, we have also examined bimolecular interactions for acetate and methoxide ions with water using the 6-31+G(d) basis set. The three-point charge TIP3P model is used to describe the solvent water.[30] Two sets of simulations were executed for comparison, one employing the standard (full SCF) combined QM/MM potential,[28] and the other utilizing the perturbative GMIPp approach.[20]

Standard procedures were used, including the isothermal−isobaric ensemble (NPT) at 298 K and 1 atm and Metropolis sampling augmented by Owicki-Scheraga preferential sampling with $1/(r^2 + C)$, where $C = 150$ Å^2. A spherical cutoff distance of 9.5 Å is used to evaluate intermolecular interactions based on heavy atom separations. For solute moves, key internal geometric parameters are varied to evaluate the effect of solvation on solute geometry. All simulations were maintained with an acceptance rate of about 45% by using ranges of ±0.15 Å and 15° for translation and rotation moves of solvent molecules, and ±0.05 to 0.15 Å and 5° to 15° for solute moves. The volume moves were kept within 130 Å^3. For each simulation, at least $(1-2) \times 10^6$ configurations of equilibration were run at the HF/3-21G QM/MM level, followed by a further equilibration of 5×10^5 configurations using HF/6-31G(d). All results are obtained by averaging over additional 2×10^6 configurations. All simulations are performed on IBM SP computers at the Minnesota Supercomputing Institute.

III. Results and Discussion

III.A. Energies. Total interaction energies and their electrostatic and van der Waals components averaged from Monte Carlo simulations using the combined QM/MM HF/6-31G(d):TIP3P and GMIPp:TIP3P potential are listed in Table 1. These quantities typically converge quickly during a simulation and, thus, can provide a good assessment of the two computational approaches in describing solute−solvent interactions. Overall, the agreement is excellent in computed ΔE_{Xs}, which is the total solute−solvent interaction energy defined by $\Delta E_{Xs} = E_{Xs}^{elec} + E_{Xs}^{vdW}$. For most molecules, the average energies from both GMIPp are in accord with those obtained using the HF QM/MM method. The difference has a root-mean-square (RMS) deviation of 1.8 kcal/mol for an energy range of −13 to −155 kcal/mol. Linear correlation analyses result in a relationship of y(HF) = 1.00 x(GMIPp) − 1.18 (kcal/mol) with $R^2 = 1.00$, where y(HF) and x(GMIPp) are energies computed using the corresponding method (Figure 1a). It is important to note that there is little deviation from unity in the correlation relationship, suggesting that there are no systematic errors between the two computational approaches and that the GMIPp/TIP3P potential can adequately reproduce interaction energies from full Hartree−Fock QM/MM calculations. Furthermore, the agreement between GMIPp:TIP3P and HF/6-31G(d):TIP3P simulations in the individual electrostatic and van der Waals terms (see Figure 1, parts b and c), demonstrates that these energy terms are also well-balanced in the perturbation approach.

Because internal coordinates are sampled during the MC simulations in solution, the geometry and conformation of the solute will be different from the optimum gas phase values because of (a) thermal fluctuations and (b) solvation effects (see below). This leads to a geometry-distortion term for the solute molecule (ΔE_X^O), which is purely due to the difference in the solute geometry in solution ($\mathbf{R}_{sol}$) and that in the gas phase ($\mathbf{R}_{gas}$). Thus

$$\Delta E_X^O = \langle \Psi^0(\mathbf{R}_{sol})|\hat{H}_{qm}|\Psi^0(\mathbf{R}_{sol})\rangle - \langle \Psi^0(\mathbf{R}_{gas})|\hat{H}_{qm}|\Psi^0(\mathbf{R}_{gas})\rangle \quad (15)$$

A comparison between the average $\langle \Delta E_X^O \rangle$ values obtained in

TABLE 1: Computed Total Solute−Solvent Interaction Energies and Their Electrostatic and van der Waals Components from QM/MM Monte Carlo Simulations in Water Using the Hartree−Fock (HF) and Generalized Molecular Interaction Potential with Polarization Correction (GMIPp) (kcal/mol)[a]

	ΔE_{Xs}^{tot}		ΔE_{Xs}^{elec}		ΔE_{Xs}^{vdW}	
solute	HF	GMIPp	HF	GMIPp	HF	GMIPp
H_2O	-14.3 ± 0.9	-14.6 ± 0.4	-14.6 ± 0.7	-15.1 ± 0.5	0.2 ± 0.2	0.6 ± 0.2
CH_3OH	-13.7 ± 0.4	-15.1 ± 0.5	-11.0 ± 0.5	-13.1 ± 0.6	-2.6 ± 0.2	-1.9 ± 0.2
CH_3OCH_3	-12.7 ± 0.3	-12.7 ± 0.4	-8.4 ± 0.2	-7.8 ± 0.4	-4.3 ± 0.1	-4.9 ± 0.1
CH_3NH_2	-16.8 ± 0.3	-13.9 ± 0.3	-15.5 ± 0.3	-12.5 ± 0.4	-1.3 ± 0.1	-1.4 ± 0.1
CH_3CN	-19.2 ± 0.5	-18.9 ± 0.5	-15.8 ± 0.5	-15.3 ± 0.5	-3.4 ± 0.2	-3.6 ± 0.2
H_2CO	-13.3 ± 0.3	-11.3 ± 0.3	-10.5 ± 0.3	-8.4 ± 0.2	-2.8 ± 0.1	-2.9 ± 0.1
$(CH_3)_2CO$	-18.9 ± 0.7	-18.9 ± 0.4	-13.5 ± 0.7	-13.7 ± 0.4	-5.4 ± 0.2	-5.2 ± 0.2
CH_3CO_2H	-22.8 ± 0.4	-19.8 ± 0.4	-19.0 ± 0.5	-15.2 ± 0.5	-3.9 ± 0.2	-4.6 ± 0.2
$HCONH_2$	-30.5 ± 0.4	-27.8 ± 0.5	-28.8 ± 0.5	-25.8 ± 0.6	-1.7 ± 0.3	-2.0 ± 0.2
NMA	-31.7 ± 0.7	-29.3 ± 0.8	-26.9 ± 0.6	-23.9 ± 0.7	-4.8 ± 0.2	-5.4 ± 0.2
pyridine	-24.7 ± 0.4	-24.2 ± 0.3	-18.4 ± 0.4	-17.3 ± 0.3	-6.3 ± 0.1	-6.9 ± 0.2
CH_3O^-	-155.0 ± 1.5	-156.3 ± 1.1	-162.6 ± 1.5	-163.8 ± 1.1	7.6 ± 0.3	7.5 ± 0.3
$CH_3CO_2^-$	-147.2 ± 0.8	-146.2 ± 1.1	-150.1 ± 0.8	-148.4 ± 1.2	2.8 ± 0.1	2.3 ± 0.2
Imidazolium	-121.9 ± 1.0	-119.0 ± 1.0	-118.4 ± 1.0	-116.5 ± 0.9	-3.5 ± 0.2	-2.5 ± 0.2

[a] The 6-31G(d) basis set is used in all calculations.

TABLE 2: Computed Distortion, Vertical Electrostatic, and Polarization Energies from HF and GMIPp QM/MM Calculations (kcal/mol)[a]

	ΔE_X^O		ΔE_{Xs}^{vert}		ΔE_{Xs}^{pol}	
solute	HF	GMIPp	HF	GMIPp	HF	GMIPp
H_2O	0.9 ± 0.1	1.0 ± 0.1	-13.7 ± 0.7	-14.4 ± 0.5	-0.9 ± 0.1	-0.8 ± 0.1
CH_3OH	2.7 ± 0.1	2.8 ± 0.1	-10.2 ± 0.5	-12.2 ± 0.6	-0.9 ± 0.1	-1.0 ± 0.1
CH_3OCH_3	4.5 ± 0.2	4.6 ± 0.1	-7.5 ± 0.2	-7.0 ± 0.4	-0.9 ± 0.1	-0.9 ± 0.1
CH_3NH_2	3.2 ± 0.1	3.1 ± 0.1	-14.1 ± 0.3	-11.7 ± 0.4	-1.4 ± 0.1	-0.8 ± 0.1
CH_3CN	2.8 ± 0.1	2.7 ± 0.2	-13.6 ± 0.5	-13.4 ± 0.5	-2.1 ± 0.2	-1.9 ± 0.1
H_2CO	1.7 ± 0.1	1.3 ± 0.1	-9.3 ± 0.3	-7.6 ± 0.2	-1.2 ± 0.1	-0.8 ± 0.1
$(CH_3)_2CO$	4.9 ± 0.1	4.9 ± 0.2	-11.6 ± 0.7	-12.1 ± 0.4	-1.9 ± 0.2	-1.7 ± 0.1
CH_3CO_2H	3.5 ± 0.1	3.6 ± 0.1	-17.3 ± 0.5	-14.1 ± 0.5	-1.7 ± 0.1	-1.1 ± 0.1
$HCONH_2$	3.1 ± 0.1	3.1 ± 0.1	-25.5 ± 0.5	-23.2 ± 0.6	-3.3 ± 0.1	-2.6 ± 0.1
NMA	6.9 ± 0.2	6.5 ± 0.2	-23.0 ± 0.6	-20.7 ± 0.7	-3.9 ± 0.2	-3.2 ± 0.2
pyridine	2.8 ± 0.1	2.7 ± 0.1	-15.3 ± 0.4	-14.4 ± 0.3	-3.2 ± 0.1	-2.9 ± 0.1
CH_3O^-	3.0 ± 0.1	3.2 ± 0.2	-158.4 ± 1.5	-159.5 ± 1.1	-4.2 ± 0.2	-4.2 ± 0.2
$CH_3CO_2^-$	4.2 ± 0.2	3.9 ± 0.1	-145.9 ± 1.1	-145.3 ± 1.2	-4.3 ± 0.1	-3.1 ± 0.1
Imidazolium	4.6 ± 0.2	5.3 ± 0.2	-117.2 ± 1.0	-115.3 ± 1.0	-1.2 ± 0.1	-1.2 ± 0.1

[a] The 6-31G(d) basis set is used in all calculations.

the GMIPp and HF simulations allows us to assess the efficiency of the GMIPp strategy to sample the solute configurational space. The agreement in the computed $\langle \Delta E_X^O \rangle$ values (Table 2) suggests that similar conformational space has been sampled when the two potential functions are used. Overall, the RMS deviation between the two sets of geometrical distortion energies is only 0.3 kcal/ mol (for a range of energies ranging from 1.0 to 5.3 kcal/mol), and as expected, there is a very good correlation between the two sets of data (y(HF) = 0.97x(GMIPp) + 0.09 (kcal/mol); $R^2 = 0.97$).

The ΔE_{Xs}^{elec} term can be further decomposed into two components, the vertical interaction energy ΔE_{Xs}^{ver} and the polarization energy ΔE_{Xs}^{pol} (eq 10).[23] The vertical interaction energy, which has the same expression in both HF and GMIPp QM/MM calculations, describes electrostatic interactions between the entire solvent system and a solute molecule with fixed gas-phase charge density. Thus, the differences between HF and GMIPp values are directly related to the difference in the conformational space of the solvent structure surrounding the solute molecule, which are sampled by the two procedures. The RMS deviation between the two methods is small: 1.7 kcal/ mol for a range of energies ranging from −7.5 to −158.4 kcal/ mol). The correlation between HF and GMIPp terms is nearly perfect ($R^2 = 1.00$), and the resulting regression equation, y(HF) = 1.00x(GMIPp) − 1.00, demonstrates that there is no systematic deviation between GMIPp and HF results. In short, both ab initio HF and GMIPp QM/MM calculations provide statistically similar results in the ΔE_{Xs}^{ver} term for the series of compounds studied here.

The way in which the solute electronic polarization energy is evaluated distinguishes the two computational procedures. In combined HF/6-31G(d):TIP3P calculations, the solute wave function is determined by SCF calculations, and thus, the polarization energy is evaluated exactly.[23] However, in the GMIPp:TIP3P approach ΔE_{Xs}^{pol} is evaluated using a second-order perturbation theory, without the need to optimize the molecular wave function.[20] Thus, a critical issue is whether such a perturbation approach can yield reasonable polarization energy in comparison with full HF−SCF calculations. We note here that the difference between the two estimated polarization energies not only reflects the intrinsic difference in the computational procedure but also includes difference in sampling in the two simulations.

For most molecules, the average GMIPp and HF polarization energies are very similar within the statistical noise of the averaging. The RMS deviation between the two computational results is about 0.5 kcal/mol in polarization energy for a range of −0.9 to −4.3 kcal/mol. The largest deviations are found for the most polarizable molecules, especially those containing carbonyl groups. Figure 2 depicts the correlation between electronic polarization energies computed using GMIPp and full HF−SCF calculations. A good linear correlation is obtained with

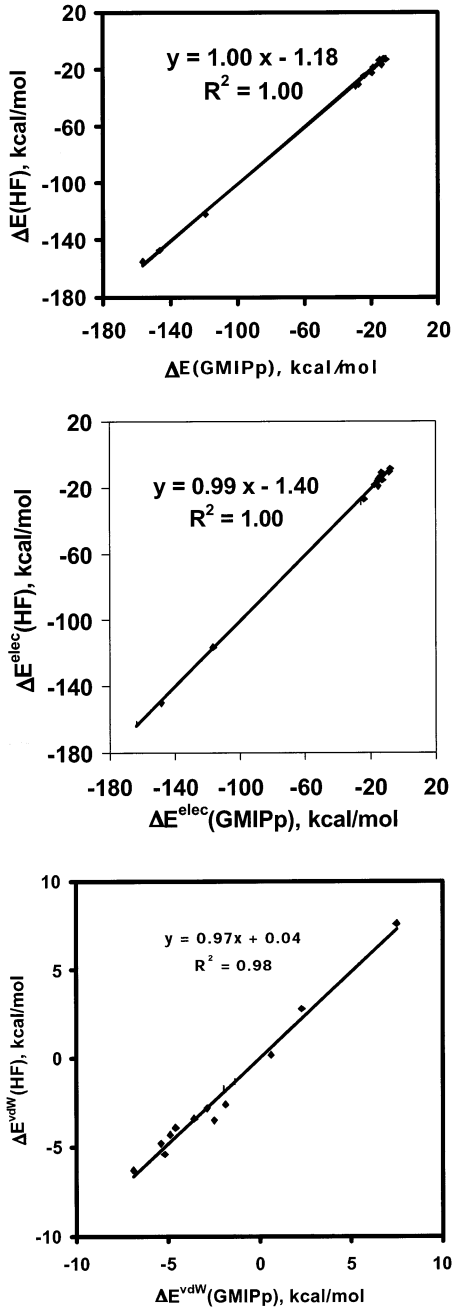

Figure 1. Correlation between energy terms computed using the full ab initio QM/MM and the GMIPp method for the total solute−solvent interaction energy (1a), the electrostatic component (1b), and the van der Waals component (1c). All results are obtained by averaging over 2 million configurations during the Monte Carlo simulation at the HF/6-31G(d) level.

$R^2 = 0.93$ and a relationship of $y(\mathrm{HF}) = 1.10x(\mathrm{GMIPp}) - 0.16$ (kcal/mol). If the small intercept is omitted, a regression equation of $y(\mathrm{HF}) = 1.16x(\mathrm{GMIPp})$ ($R^2 = 0.92$) is obtained. The deviation from unity in the slope suggests that the perturbation approach underestimates polarization effects by about 10−16%, which may be attributed to the neglect of higher order perturbation terms. This finding is in good accord with previous findings in the study of hydrogen-bonded complexes, where the perturbation approach was found to consistently underestimate polarization energy by about 10%.[31]

For anionic species, it is typically necessary to include diffuse functions to adequately describe their reactivity. Although the present study is aimed at examining the performance of a perturbation approach to the calculation of solute−solvent interaction energies in Monte Carlo simulations, it is interesting to investigate the stability of the GMIPp method when diffuse functions are included, which could bring in more low-lying virtual orbitals. Table 3 compares the computed hydrogen bonding energies using HF/6-31+G(d) and GMIPp/6-31+G-(d) methods for bimolecular interactions of acetate and methoxide ions with one or two water molecules in three different configurations (Scheme 1). In all calculations, the geometries of the isolated monomers were used in the complex and the water molecule was placed 3.0 Å from anion for O−O atom separations. The TIP3P model was used for water in GMIPp calculations. Table 3 shows that the agreement between interaction energies from explicit HF calculations and the perturbation approach is excellent. This indicates that the present GMIPp method is stable when diffuse functions are included for anionic species.

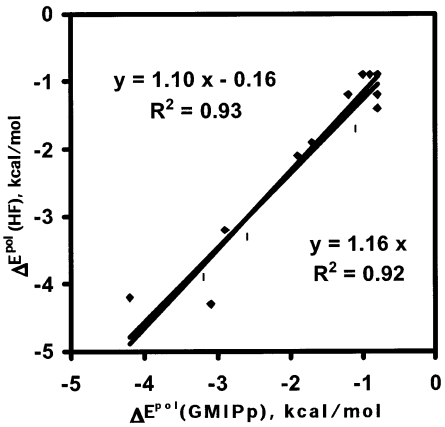

Figure 2. Correlation between polarization energies determined using the ab initio QM/MM and GMIPp method. All results are obtained by averaging over 2 million configurations during the Monte Carlo simulation at the HF/6-31G(d) level.

SCHEME 1

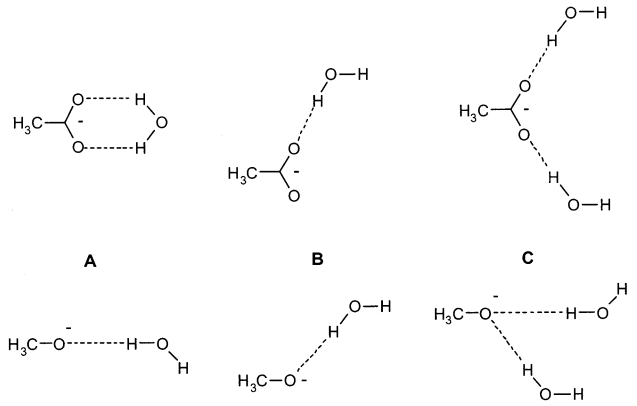

TABLE 3: Polarization Energies (kcal/mol) Computed at the Hartree−Fock and GMIPp Levels of Theory Using the 6-31+G(d) Basis Set for Hydrogen-Bonding Complexes of Methoxide and Acetate Ions with Water

	CH_3O^-		CH_3COO^-	
complex	HF	GMIPp	HF	GMIPp
A	−0.4	−0.3	−0.5	−0.4
B	−0.4	−0.3	−0.4	−0.3
C	−1.2	−1.0	−0.6	−0.5

III.B. Induced Dipole Moment. Solvation alters a solute's electrostatic properties by changing its geometry and by polarizing its charge distribution. The polarization effect by the solvent is reflected by the change in molecular dipole moment

TABLE 4: Computed Dipole Moments (Debye) in the Gas Phase and in Aqueous Solution from Combined QM/MM HF/6-31G(d) Simulations

solute	μ (gas)	μ (exp)	$\langle\mu\rangle$ (aq)	$\Delta\mu_{ind}$	$\Delta\mu_{ind}$ (AM1)[a]
H_2O	2.20	1.85	2.50 ± 0.02	0.30	0.29
CH_3OH	1.87	1.70	2.23 ± 0.02	0.35	0.44
CH_3OCH_3	1.48	1.30	1.98 ± 0.02	0.50	0.44
CH_3NH_2	1.53	1.31	2.02 ± 0.02	0.49	0.20
CH_3CN	4.04	3.92	5.09 ± 0.04	1.05	0.85
H_2CO	2.67	2.32	3.29 ± 0.02	0.62	
$(CH_3)_2CO$	3.12	2.88	4.16 ± 0.06	1.04	0.95
CH_3CO_2H	1.80	1.74	2.57 ± 0.02	0.77	0.32
$HCONH_2$	4.10	3.73	5.54 ± 0.03	1.44	
NMA	3.94	3.85	5.77 ± 0.04	1.83	1.67
pyridine	2.31	2.22	3.67 ± 0.05	1.32	

[a] Reference 23.

of the solute on going from the gas phase into solution. Such a change can be easily determined once the solute wave function in solution is obtained. This can be done using either the HF QM/MM potential or the GMIPp perturbation approach, although the later strategy has not been tested here.

Table 4 lists the HF/6-31G(d) dipole moments for all molecules in the gas phase and in aqueous solution except ionic species. The gas-phase dipole moments can be compared with the available experimental data, but there are no experimental data for solution phase dipole moments. Previously, we have carried out combined QM/MM simulations of a similar set of compounds using the semiempirical AM1 (Austin model 1) method, and the results from that work are also listed in Table 4 for comparison.[23] Although Hartree−Fock theory overestimates gas-phase dipole moments, its change up on solvation, i.e., the induced dipole moment, can be reasonably estimated using HF/6-31G(d) in combined QM/MM simulations. The aqueous solvation effect leads to a systematic increase in the dipole moment by about 32%, which is similar to that reported in previous ab initio and semiempirical SCRF calculations.[32,33] Interestingly, there is also a general good agreement between QM/MM estimates of the induced dipole obtained from HF/6-31G(d) and AM1 calculations (see Table 4 and reference 23).

III.C. Geometry. The transfer of a solute molecule from gas phase into solution leads to variations in its molecular geometry. During our Monte Carlo simulation, geometries of the solute molecules were included in the Metropolis sampling. Thus, we have obtained average solute geometries for the solute molecules in aqueous solution. Overall, the RMS deviation between computed bond distances in the gas phase and in solution is 0.011 Å. Figure 3 shows that except for C−H bonds, which are shortened upon hydration, there is a general tendency to increase the chemical bonds upon hydration. This effect is especially noticeable for bonds between atoms engaged in hydrogen bonding interactions with the solvent. It is particularly interesting to note that formamide and acetamide have a much elongated C═O bond length, which is accompanied by a concomitant decrease in the C−N bond. This is consistent with the large increase in molecular dipole moments for the amides (Table 4), where solvation enhances amide conjugation. Furthermore, the rotational barrier about the amide bond is known to have large solvent effects and increases as the solvent polarity increases.[34−37] This is a reflection of the more pronounced double bond character between the carbonyl carbon and amide nitrogen.

Solvent effects have little impact on bond angles for the compounds studied here, with an average RMS deviation of about 1° between the gas phase and solution values. Furthermore, the effect of hydration in bond angles is less uniform

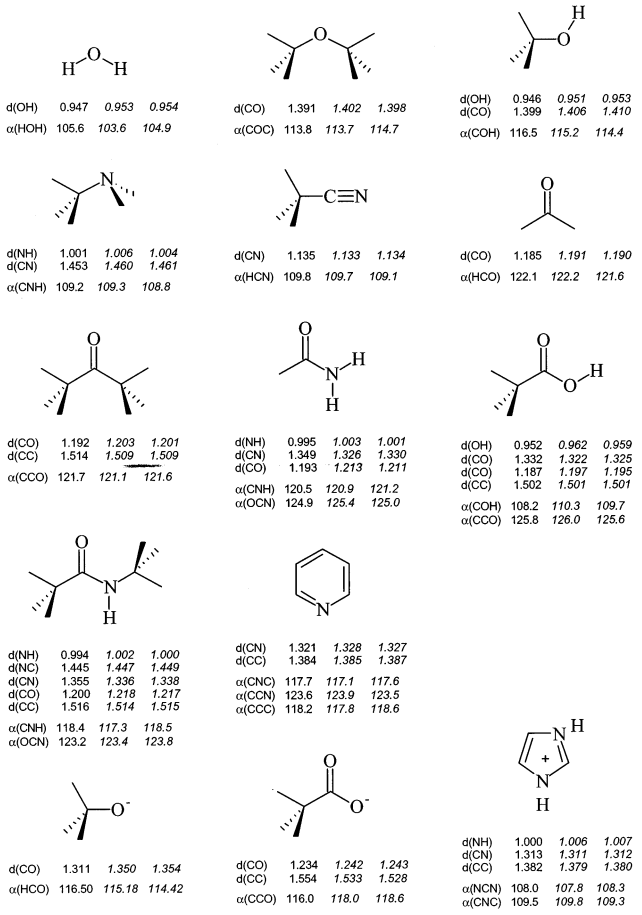

Figure 3. Computed molecular geometries in the gas phase and in solution. In each structure, the gas-phase values are followed by results from full ab initio QM/MM simulation and those from the GMIPp calculations, respectively. Bond distances are in angstroms, and bond angles are in degrees.

(see Figure 3), and it clearly depends on the solute molecule. Overall, the agreement between HF and GMIPp calculations is excellent. Bond lengths and angles in solution are reproduced with a RMS error of just 0.002 Å and 0.8°, respectively, between the two computational methods. Thus, the GMIPp strategy is able to provide a good estimate of the solute geometry in aqueous solution.

III.D. Radial Distribution Functions. Analysis of the solute−solvent radial distribution functions allows us to gain insight into the structure of the solvent around the solute molecule on a statistical basis. Some of the computed radial distribution functions from the simulation of acetate ion, formamide, acetonitrile, and pyridine in water are shown in Figure 4, which illustrate the agreement between the full HF QM/MM and the present GMIPp QM/MM simulations. The location of all $X-O_W$ radial distribution functions, where X is a non-hydrogen atom in the solute and O_W is water oxygen, are within 0.1 Å, and the peak heights differ by about 0.2 on average. These results indicate that the perturbation approach is a promising alternative for combined QM/MM simulations, which yield adequate energetic and structural data in comparison with explicit HF QM/MM calculations.

III.E. Computational Efficiency. An obvious advantage of the GMIPp method is that it avoids the repeated SCF calculations of the solute electronic structure problem during the fluid simulation. This reduces the computational costs, potentially enabling accurate ab initio methods to be applied to large molecular systems in solution phase and biomolecular simula-

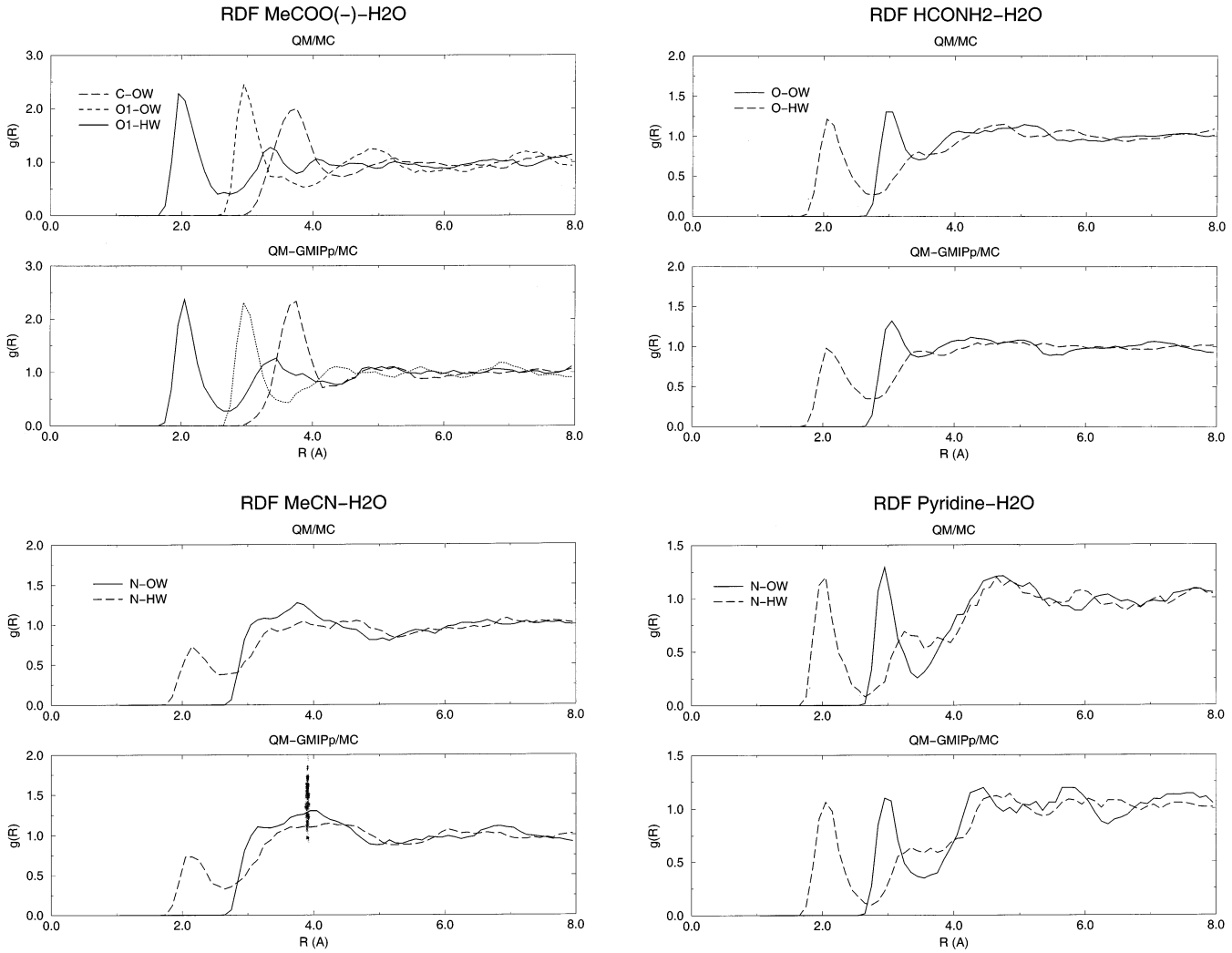

Figure 4. Computed radial distribution functions for acetate ion (4a), formamide (4b), acetonitrile (4c), and pyridine (4d) in water.

tions. For the molecules studied here, consisting of 19−100 basis functions, the computational costs are reduced by 36% for water (19 basis functions) to more than 65% for pyridine (100 basis functions). Interestingly, the representation of the ratio between the CPU time of the HF and GMIPp calculations shows a linear relationship ($R = 0.92$), suggesting that larger systems gain greater computational savings using the GMIPp method.

We note that the present test simulation of the GMIPp method has been carried out conservatively because it is possible to further reduce the number of virtual orbitals used in the perturbation calculation. In addition, we have repeated the SCF calculation for the "gas phase" solute molecule in every Monte Carlo movement involving a volume change and a solute move. If the solute internal geometry is not varied, one can use a simple transformation to obtain the new orbital coefficients and electronic integrals in the new solute orientation, avoiding additional SCF calculations. Thus, in principle, one only needs to do a single SCF calculation of the solute molecule in the gas phase at the beginning of combined QM/MM fluid simulations in the GMIPp approach.

IV. Conclusions

By comparison with results obtained from full ab initio HF/6-31G(d) QM/MM simulations, we have demonstrated that the use of the generalized molecular interaction potential with polarization correction (GMIPp) gives an accurate representation of solute−solvent interactions. Monte Carlo QM/MM calculations using the GMIPp potential and the full HF QM/MM method yield similar energetic and structural results for a series of 14 organic compounds in aqueous solution. The solute electronic polarization energy is reasonably estimated using the GMIPp method.

The implementation of the GMIPp within the Metropolis-Monte Carlo framework makes possible extensive QM/MM calculations of large molecules, which otherwise would be difficult using standard HF QM/MM methods. At this point, it is worth noting that previous studies have shown that the GMIPp potential is reliable for describing the polarization energy in complex systems involving a variety of molecular species (charged and neutral) and interactions, such as salt bridges,[20] hydrogen-bonded complexes, [25] cation−π interactions,[26] and the stacking between heterocyclic compounds.[31] For the molecules studied here, we used a split valence basis set plus polarization functions in both GMIPp and HF calculations, and the former method reduces the total cost of the fluid simulations by a factor from 1.6 to almost 3 times, without apparent loss of accuracy, despite the conservative protocol used in this study. Moreover, it has been shown that the polarization energies computed from the GMIPp potential can be scaled to reproduce the values determined using theories that include explicit electron correlation effects.[38] These encouraging results allow us to anticipate

that the GMIPp method should be useful for ab initio QM/MM studies of biomolecular complexes including enzymatic reactions.

Acknowledgment. This work has been supported by the Spanish DGICYT (PM99-0046 and PB98-1222), by NATO support for USA-Spain collaborative projects (CRG 972068), and by the National Institutes of Health (J.G.) and National Science Foundation (J.G.). E.C. thanks a fellowship from the Universitat de Barcelona.

References and Notes

(1) Gao, J. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH: New York, 1995; Vol. 7, pp 119−185.
(2) Gao, J. *Acc. Chem. Res.* **1996**, *29*, 298−305.
(3) Field, M. J.; Bash, P., A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700−733.
(4) Alhambra, C.; Corchado, J.; Sanchez, M. L.; Gao, J.; Truhlar, D. G. *J. Am. Chem. Soc.* **2000**, *122*, 8197−8203.
(5) Rothlisberger, U.; Carloni, P.; Doclo, K.; Parrinello, M. *J. Biol. Inorg. Chem.* **2000**, *5*, 236−250.
(6) Cui, Q.; Karplus, M. *J. Am. Chem. Soc.* **2002**, *124*, 3093−3124.
(7) Gao, J.; Truhlar, D. G. *Ann. Rev. Phys. Chem.* **2002**, *53*, 467−505.
(8) Truhlar, D. G.; Gao, J.; Alhambra, C.; Garcia-Viloca, M.; Corchado, J.; Sanchez, M. L.; Villa, J. *Acc. Chem. Res.* **2002**, ACS ASAP.
(9) Mo, Y.; Gao, J. *J. Comput. Chem.* **2000**, *21*, 1458−1469.
(10) Mo, Y.; Gao, J. *J. Phys. Chem. A* **2000**, *104*, 3012−3020.
(11) Eichinger, M.; Tavan, P.; Hutter, J.; Parrinello, M. *J. Chem. Phys.* **1999**, *110*, 10452−10467.
(12) Gao, J. *J. Am. Chem. Soc.* **1995**, *117*, 8600−8607.
(13) Evans, T. J.; Truong, T. N. *J. Comput. Chem.* **1998**, *19*, 1632−1638.
(14) Truong, T. N.; Stefanovich, E. V. *Chem. Phys. Lett.* **1996**, *256*, 348−352.
(15) Devi-Kesavan, L. S.; Garcia-Viloca, M.; Gao, J. *Theor. Chem. Acc.* **2003**, in press.
(16) Cui, Q.; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Karplus, M. *J. Phys. Chem. B* **2001**, *105*, 569−585.
(17) Freindorf, M.; Gao, J. *J. Comput. Chem.* **1996**, *17*, 386-395.
(18) Orozco, M.; Luque, F. J. *J. Comput. Chem.* **1993**, *14*, 587−602.
(19) Alhambra, C.; Luque, F. J.; Orozco, M. *J. Phys. Chem.* **1995**, *99*, 3084−3092.
(20) Luque, F. J.; Orozco, M. *J. Comput. Chem.* **1998**, *19*, 866−881.
(21) Thompson, M. A. *J. Phys. Chem.* **1996**, *100*, 14492−14507.
(22) Gao, J. *J. Comput. Chem.* **1997**, *18*, 1062−1971.
(23) Gao, J.; Xia, X. *Science* **1992**, *258*, 631−635.
(24) Gao, J.; Luque, F. J.; Orozco, M. *J. Chem. Phys.* **1993**, *98*, 2975−2982.
(25) Hernandez, B.; Luque, F. J.; Orozco, M. *J. Comput. Chem.* **1999**, *20*, 937−946.
(26) Cubero, E.; Luque, F. J.; Orozco, M. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 5976−5980.
(27) Gao, J. 4.0 ed.; Department of Chemistry, University of Minnesota: Minneapolis, MN, 2000.
(28) Gao, J.; Freindorf, M. *J. Phys. Chem. A* **1997**, *101*, 3182−3188.
(29) 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. S.; 11 ed., 1993.
(30) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(31) Luque, F. J.; Lopez, J. M.; Lopez de la Paz, M.; Vicent, C.; Orozco, M. *J. Phys. Chem. A* **1998**, *102*, 6690−6696.
(32) Cramer, C. J.; Truhlar, D. G. *Science* **1992**, *256*, 213−217.
(33) Cramer, C. J.; Truhlar, D. G. *Chem. Phys. Lett.* **1992**, *198*, 74−80.
(34) Duffy, E. M.; Severance, D. L.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1992**, *114*, 7535−7542.
(35) Gao, J. *J. Am. Chem. Soc.* **1993**, *115*, 2930−2935.
(36) Luque, F. J.; Orozco, M. *J. Org. Chem.* **1993**, *58*, 6397−6405.
(37) Luque, F. J.; Orozco, M. *J. Chem. Soc., Perkin Trans. 2* **1993**, 683−690.
(38) Chipot, C.; Luque, F. J. *Chem. Phys. Lett.* **2000**, *332*, 190.