JCTC Journal of Chemical Theory and Computation

Polarizable Simulations with Second-Order Interaction Model—Force Field and Software for Fast Polarizable Calculations: Parameters for Small Model Systems and Free Energy Calculations

George A. Kaminski,*,† Sergei Y. Ponomarev,† and Aibing B. Liu†,‡

Department of Chemistry and Biochemistry, Worcester Polytechnic Institute, Worcester, Massachusetts 01609 and Department of Chemistry, Central Michigan University, Mt. Pleasant, Michigan 48859

Received August 6, 2009

Abstract: We are presenting POSSIM (POlarizable Simulations with Second-order Interaction Model)—a software package and a set of parameters designed for molecular simulations. The key feature of POSSIM is that the electrostatic polarization is taken into account using a previously introduced fast formalism. This permits cutting the computational cost of using the explicit polarization by about an order of magnitude. In this article, parameters for water, methane, ethane, propane, butane, methanol, and N-methylacetamide (NMA) are introduced. These molecules are viewed as model systems for protein simulations. We have achieved our goal of ca. 0.5 kcal/mol accuracy for gas-phase dimerization energies and no more than 2% deviations in liquid state heats of vaporization and densities. Moreover, free energies of hydration of the polarizable methane, ethane, and methanol have been calculated using the statistical perturbation theory. These calculations serve as a model for calculating protein pK_a shifts and ligand binding affinities. The free energies of hydration were found to be 2.12, 1.80, and −4.95 kcal/mol for methane, ethane, and methanol, respectively. The experimentally determined literature values are 1.91, 1.83, and −5.11 kcal/mol. The POSSIM average error in these absolute free energies of hydration is only about 0.13 kcal/mol. Use of the statistical perturbation theory with polarizable force fields is not widespread, and we believe that this work opens the road to further development of the POSSIM force field and its applications for obtaining accurate energies in protein-related computer modeling.

I. Introduction

Computer simulations have become an integral part of chemical and biochemical research, delivering answers to a variety of questions ranging from assessing reaction kinetic data to providing microscopic insight into systems involving proteins and DNA. The key issue in any such modeling is the way the energy of the system is evaluated. Quantum mechanics provides a robust tool for this task, but there are two general problems. First, quantum mechanical calculations are resource-demanding; therefore, the size of a system which can be treated this way is limited. Second, there is no one single recipe for obtaining uniformly accurate quantum mechanical data, and the level of theory required for different problems varies.

Empirical force fields offer an alternative which is much cheaper computationally. However, it has been shown that, in many cases, accurate assessment of energy with empirical force fields necessitates explicit treatment of many-body interactions, mainly the electrostatic polarization.[1] Dimerization energies and acidity constants of small molecules, energies of protein−ligand interactions, protein pK_a values, or even the very thermodynamic stability of complexes in

* Corresponding author e-mail: gkaminski@wpi.edu.
† Worcester Polytechnic Institute.
‡ Central Michigan University.

10.1021/ct900409p CCC: $40.75
Published on Web 10/05/2009

solutions often depend on including the electrostatic polarization into the empirical Hamiltonian. For example, it has been demonstrated in earlier works that pK_a values for acidic and basic residues of the OMTKY3 can be reproduced within 0.6 and 0.7 pH units of the experimental data if explicit treatment of polarization is included. The errors with the nonpolarizable OPLS are 3.3 pH units for the acidic residues and 2.2 for the basic ones.[2] Another interesting example was a study of Sialyl LewisX in complex with SelectinE. This sugar−protein complex is known to exist experimentally but dissociated in molecular dynamics simulations with fixed charges.[3] Overall, many-body interactions play a crucial role in many applications, although they are sometimes included in surrogate forms, for example, as conformation-specific protein charges.[4]

There are two major issues pertaining to polarizable calculations which are still not uniformly resolved. On one hand, the optimal source of fitting data seems to vary from application to application. While it is attractive to rely almost solely on high-level quantum mechanical results,[5] experimental data often present more robustness. Moreover, while quantum mechanical calculations permit a great level of microscopic insight, the very values of quantum mechanical energies are often uncertain, as they may significantly change depending on the level of theory. We adopt a middle-path approach in this work, in which we rely on experimental data whenever possible and make heavy use of quantum mechanical calculations, while trying not to treat them as panacea.

The second issue with polarizable force fields is the functional form of polarization itself. There are several viable techniques present, including fluctuating charges or inducible point dipoles. It is possible to implement these quite efficiently for uniform systems known in advance (such as pure water[6]). But software for simulating arbitrary polarizable systems (including proteins and protein−ligand complexes) is usually significantly slower than its fixed-charges counterparts. Calculating induced dipoles or fluctuating charges requires solving a system of self-consistent equations. This is usually done iteratively, and the process may lead to unlimited growth of the induced dipoles (the so-called polarization catastrophe). In order to reduce the computational cost of polarizable calculations, we introduced an approximation that we termed second-order polarization. It permitted a reduction of the time needed for assessing the polarization energy by about an order of magnitude, removed any possibility of polarization catastrophe, and was done without any sacrifice of the computational accuracy.[7]

This article reports the next stage of the systematic development of the second-order polarization technique. A software suite called POSSIM (POlarizable Simulations with Second-order Interaction Model) has been created for the second-order polarizable calculations. Parameters for several model systems have been developed. The list of these systems includes N-methylacetamide (NMA), which will be further used as the main building block in protein backbones. Moreover, free energy perturbations were performed with the POSSIM force field and software to obtain relative and absolute Gibbs free energies of hydration for methane, ethane, and methanol. While being widely used with fixed-charges force fields, such calculations are still rather rare with polarizable techniques. The toolset presented in this paper will be further utilized in developing POSSIM parameters for proteins and other systems and studying biophysical and organic processes, including protein−ligand binding.

In this work, we benchmark the POSSIM results (marked PFF, for Polarizable Force Field) against the fixed-charges OPLS-AA[8] as well as the previous version of the polarizable force field (non-second-order, thus a slower one) termed PFF0.[9]

The rest of the paper is organized as follows: Given in section II is a description of the methodology involved. Section III contains results and discussion. Finally, conclusions are presented in section IV.

II. Methods

A. Force Field. The total energy E_{tot} is calculated as follows:

$$E_{tot} = E_{electrostatic} + E_{vdW} + E_{stretch} + E_{bend} + E_{torsion} \quad (1)$$

$E_{electrostatic}$ stands for the electrostatic interactions, including the dipole−dipole, dipole−charge, and charge−charge contributions. E_{vdW} is the nonelectrostatic part of the nonbonded inter- and intramolecular energy; $E_{stretch}$ and E_{bend} are the harmonic bond stretching and angle bending, respectively. Finally, $E_{torsion}$ is the Fourier expansion for the torsional energy.

Electrostatic Energy. The electrostatic polarization energy with inducible point dipoles is

$$E_{pol} = -\frac{1}{2}\sum_i \mu_i E_i^0 \quad (2)$$

E^0 denotes the electrostatic field in the absence of the dipoles, and μ represents the induced dipole moments, which are calculated as follows:

$$\mu_i = \alpha_i E_i^{tot} \quad (3)$$

where α represents scalar polarizabilities and E_i^{tot} is the total field, including the dipole−dipole component:

$$E_i^{tot} = E_i^0 + \sum_{j\neq i} T_{ij}\mu_j \quad (4)$$

$$T_{ij} = \frac{1}{R_{ij}^3}\left(\frac{3R_{ij}R_{ij}}{R_{ij}^2} - \mathbf{I}\right) \quad (5)$$

with R_{ij} being the distances between atomic sites i and j. $\mathbf{I}$ is the unit tensor. Then,

$$\mu_i = \alpha_i E_i^0 + \alpha_i \sum_{j\neq i} T_{ij}\mu_j \quad (6a)$$

or

$$\mathbf{A}\mu = E^0 \quad (6b)$$

The self-consistent eq 6 is usually solved iteratively. If we explicitly write down the first three iterations, we can produce "first-order", "second-order", and "third-order" approximations for the induced dipoles:

$$\mu_i^0 = \alpha_i E_i^0 \tag{7a}$$

$$\mu_i^{\mathrm{I}} = \alpha_i E_i^0 + \alpha_i \sum_{j \neq i} T_{ij} \mu_j^0 = \alpha_i E_i^0 + \alpha_i \sum_{j \neq i} T_{ij} \alpha_j E_j^0 \tag{7b}$$

$$\mu_i^{\mathrm{II}} = \alpha_i E_i^0 + \alpha_i \sum_{j \neq i} T_{ij} \mu_j^I = \alpha_i E_i^0 + \alpha_i \sum_{j \neq i} T_{ij} \alpha_j E_j^0 + \alpha_i \sum_{j \neq i} T_{ij} \alpha_j \sum_{k \neq j} T_{jk} \alpha_k E_k^0 \tag{7c}$$

The first-order expression requires considerably less resources than the exact expression in eq 6, as the matrix **A** does not come into play, but dipole−dipole interactions are totally ignored. It has been shown that, although more accurate than the fixed-charges description, this technique is not accurate enough in describing molecular systems.[10] We are utilizing the second-order expression in eq 7b instead. It has been previously shown to yield about an order of magnitude increase of the computational speed with no loss of accuracy.[7]

The overall electrostatic energy

$$E_{\mathrm{electrostatic}} = E_{\mathrm{pol}} + E_{\mathrm{additive}} \tag{8}$$

where

$$E_{\mathrm{additive}} = \sum_{i \neq j} \frac{q_i q_j}{R_{ij}} f_{ij} \tag{9}$$

represents the pairwise-additive Coulomb charge−charge interaction energies between charges on atoms i and j. The term f_{ij} equals zero for 1,2 and 1,3 pairs (atoms which belong to the same valence bond or angle), 0.5 for 1,4 interactions (atoms in the same dihedral angle), and 1.0 for all of the other pairs.

To avoid a nonphysical increase of the electrostatic interactions at close distances, each atom is assigned a cutoff parameter, R_{cut}. When the overall distance R_{ij} is smaller than the sum of these parameters $R_{\mathrm{min}}{}^{ij} = R_{\mathrm{cut}}{}^{i} + R_{\mathrm{cut}}{}^{j}$ for the atoms i and j, R_{ij} is replaced by an effective smooth function

$$R_{ij}^{\mathrm{eff}} = \left[1 - \left(\frac{R_{ij}}{R_{\mathrm{min}}^{ij}}\right)^2 + \left(\frac{R_{ij}}{R_{\mathrm{min}}^{ij}}\right)^3\right] \times R_{\mathrm{min}}^{ij} \tag{10}$$

The Rest of the Force Field. The van der Waals energy is evaluated with the standard Lennard-Jones formalism:

$$E_{\mathrm{vdW}} = \sum_{i \neq j} 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{R_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{R_{ij}}\right)^{6}\right] f_{ij} \tag{11}$$

Geometric combining rules are used:

$$\varepsilon_{ij} = (\varepsilon_i \cdot \varepsilon_j)^{1/2}, \sigma_{ij} = (\sigma_i \cdot \sigma_j)^{1/2} \tag{12}$$

and the coefficient f_{ij} is calculated the same way as for the electrostatic term.

The bond-stretching and angle-bending energies were obtained in accordance with eqs 13 and 14.

$$E_{\mathrm{stretch}} = \sum_{\mathrm{bonds}} K_r (r - r_{\mathrm{eq}})^2 \tag{13}$$

$$E_{\mathrm{bend}} = \sum_{\mathrm{angles}} K_\Theta (\Theta - \Theta_{\mathrm{eq}})^2 \tag{14}$$

Here, the subscripts eq are used to denote the equilibrium values of the bond length r and angle Θ.

Finally, the torsional term was computed as follows:

$$E_{\mathrm{torsion}} = \sum_i \frac{V_1^i}{2}[1 + \cos(\phi_i)] + \frac{V_2^i}{2}[1 - \cos(2\phi_i)] + \frac{V_3^i}{2}[1 + \cos(3\phi_i)] \tag{15}$$

with the summation performed over all of the dihedral angles i.

When a comparison is done with results obtained with the nonpolarizable OPLS-AA force field,[8] the latter lacks the first (polarizable) term in the electrostatic energy expression in eq 8. The previous generation polarizable force field[9] (denoted as PFF0) employs the full iterative solvation of eq 6, has permanent dipole moments in addition to the inducible ones, and uses the exp-6 van der Waals energy term instead of the Lennard-Jones 12−6 formalism of this work.

B. Parameterization of the Force Field. The procedure for determining values of the potential energy parameters consisted of the following stages: (i) fitting the electrostatic polarizabilities, (ii) fitting the permanent electrostatic charges, (iii) determining Lennard-Jones parameters, (iv) obtaining values of the torsional parameters, and (v) fine-tuning of the force field. All of the calculations for the newly developed PFF (both for the parametrization and free energy perturbations) were performed with the POSSIM software suite.

Electrostatic Polarizabilities. A series of electrostatic perturbations was applied to the target molecules, in the form of dipolar probes consisting of two opposite charges of magnitude 0.78e, 0.58 Å apart (for a dipole moment of 2.17 D—similar to that of nonpolarizable models for liquid water such as SPC/E[11]), placed at locations where hydrogen bonds to the molecule were formed. The perturbations were the same as used for the previous generation of the polarizable force field (PFF0).[5,12] For each perturbation, the change in the electrostatic potential at a set of gridpoints outside the van der Waals surface of the molecule, as well as the energy of the perturbed system, was computed using density-functional theory (DFT) with the B3LYP method[13] and cc-pVTZ(-f) basis set. All calculations were performed with the Jaguar electronic structure code.[14] The polarizablities α_i are assumed to be isotropic and are chosen to minimize the deviation of the three-body energy obtained with the PFF and the DFT calculations. The three-body energies were calculated in accordance with eq 16 and Figure 1.

$$E_{\mathrm{3body}} = E(1+2+3) - E(1+2) - E(1+3) - E(2+3) + E(1) + E(2) + E(3) \tag{16}$$

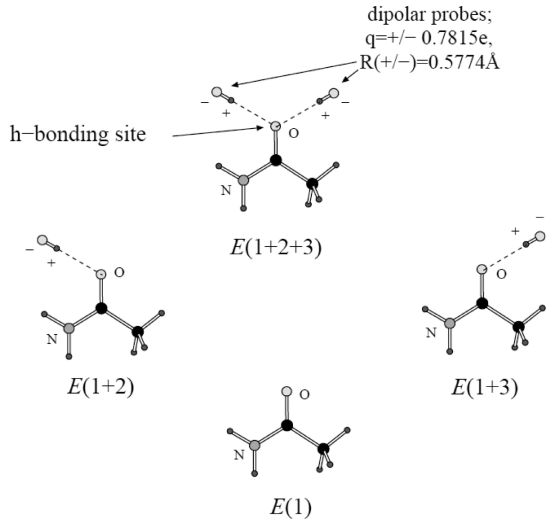

Figure 1. Calculating two- and three-body energies of a molecule with dipolar probes.

Here, the molecule for which parameters are produced is denoted as 1, and the two dipolar probes are marked as 2 and 3. This three-body energy is automatically equal to zero in a nonpolarized fixed-charges force field, in which no many-body interactions are explicitly present. It has been shown that the three-body energy is independent of the permanent charges and depends only on the values of the polarizabilities.[7] Therefore, it is justified that the polarizabilities in our force field were parametrized first.

The choice of electronic structure method (DFT/B3LYP functional, cc-pVTZ(-f) basis set) yields quite accurate permanent charge distributions but underestimates the gas-phase polarizability as compared to experimental results. Closer agreement with gas-phase experiments could be obtained by including diffuse functions in the DFT calculations. However, our previous computational experiments with liquid-state simulations strongly suggest that these diffuse function contributions are considerably damped in the condensed phase, and that ignoring them is in fact a much better approximation than fully including them.[5] Briefly, in the condensed phase, Pauli repulsion from neighboring molecules raises the energies of diffuse functions and so diminishes their contribution to the polarization. Empirically, when diffuse functions are used to develop polarization responses for small molecules, liquid-state simulations of these molecules manifest overpolarization of the solvent.

Permanent Electrostatic Charges. We have used the same quantum mechanical systems as above, except that two-body energies were employed as the fitting target:

$$E_{2\text{body}} = E(1 + 2) - E(1) - E(2) \tag{17}$$

The polarizabilities were not changed from their values obtained at the previous step. The charges were adjusted to minimize the two-body rms deviations.

In all of the cases (except for the dipolar probes), the R_{cut} for both dipoles and charges on the molecule involved were set to 0.8 Å.

Lennard-Jones Parameters. Fitting the Lennard-Jones part of the force field was done with high-accuracy ab initio results for intermolecular hydrogen-bonding interaction energies and distances as a target. Gas-phase energy minimizations were carried out. The quantum mechanical data were obtained as described in ref 15, and many actual target energies were adopted from the refs 9 and 12 (as noted in the tables in the Results and Discussion section). The methods employed to calculate binding energies are based on an MP2 extrapolation procedure that was previously developed using the pseudospectral local MP2 (LMP2) approach.

The goal was to reproduce the gas-phase intermolecular binding affinities and geometries as accurately as possible. We normally set a target of 0.25−0.5 kcal/mol or better for the precision of the binding affinity. For hydrogen bonds, this can be achieved via MP2 calculations extrapolated to the basis set limit, where the contribution of higher-level excitations (e.g., CCSD(T)) has been shown to be negligible (although, in some cases, such as π stacking of aromatic rings, the MP2 level is not adequate to achieve the target accuracy).

Briefly, dimer geometries were obtained by LMP2 optimizations with a cc-pVTZ(-f) basis set.[16] The empirical dimer binding energy consists of the LMP2 binding energy for a smaller cc-pVTZ(-f) basis set (E_{ccpvtz}) and the LMP2 binding energy with a larger cc-pVQZ(-g) basis set (E_{ccpvq}). The model binding energy E_{bind} takes the simple form[15]

$$E_{\text{bind}} = C_1 \cdot E_{\text{ccpvtz}} + C_2 \cdot E_{\text{ccpvqz}} \tag{18a}$$

$$C_1 = a_1/(a_1 - a_2); C_2 = -a_2/(a_1 - a_2) \tag{18b}$$

$$a_1 = \exp(-2.7); a_2 = \exp(-1.8) \tag{18c}$$

In calculating binding energies, the Hartree−Fock energies are corrected for basis set superposition error using the counterpoise method.

The target hydrogen-bonded distances were taken directly from the LMP2/cc-pVTZ(-f) energy minimizations.

Bond Stretching and Angle Bending. The bond-stretching and angle-bending energies were obtained according to eqs 13 and 14. The values of the parameters were taken directly from the OPLS-AA.[8] The reason is that both OPLS-AA and our PFF disregard 1,2 (covalent bond) and 1,3 (covalent angle) interactions; therefore, the energetics related to these arrangements are the same with both techniques.

Torsional Parameters. Finally, the torsional term was computed as shown in eq 15. We performed constrained geometry optimizations with dihedral angles fixed at their key positions and optimized the parameters V to minimize deviations of the relative energies from the LMP2/cc-pVTZ(-f) quantum mechanical results. For instance, for the ethane molecule, constraining the H−C−C−H dihedral to 60° and to 0° yielded relative quantum mechanical energies of 0 and 2.7762 kcal/mol, respectively. We have optimized the V_3 term of this torsion to achieve a close agreement with these numbers (0 and 2.7751 kcal/mol). The V_1 and V_2 terms in this case were set to zero.

Liquid Simulations. The final tuning of the force field parameters was achieved by reproducing experimental values of heats of vaporization and molecular volumes of the pure

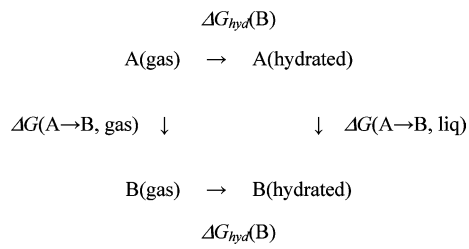

Figure 2. Thermodynamic cycle used to assess relative hydration energies.

molecular liquids involved. Each simulation was run with the POSSIM software and included 216 molecules in a cubic cell with periodic boundary conditions. The NPT ensemble (constant temperature, pressure, and the number of molecules) was employed. Methanol and water were simulated at 25 °C. The NMA liquid had a temperature of 100 °C. The hydrocarbons were modeled at their boiling temperatures: −161.49 °C for methane, −88.63 °C for ethane, −42.1 °C for propane, and −0.5 °C for butane. The calculations were carried out with the Monte Carlo technique, and the heats of vaporization were calculated according to eq 19:

$$\Delta H_{\mathrm{vap}} = E(\mathrm{gas}) - E(\mathrm{liq}) + RT \qquad (19)$$

The difference between the energy for one molecule in the gas-phase and in the condensed state was augmented by the RT term to account for the $\Delta(PV)$ part of the enthalpy, in the assumption that the vapor obeys the ideal gas law, and the molecular volume of the liquid can be neglected compared to that of the gas. In all of the calculations, at least 1×10^6 Monte Carlo configurations of averaging were followed by no less than 5×10^6 configurations of averaging for the thermodynamic properties. Elements of the dipole−dipole interaction tensor in eq 5 were set to zero for distances beyond 7.0 Å. The other intermolecular interactions were cut off at 8.0 Å for water; 10.0 Å for methane and methanol; and 11.0 Å for ethane, propane, butane, and NMA. The charge−charge interactions were switched off smoothly over the last 0.5 Å. The standard correction for the neglected Lennard-Jones energies beyond the cutoff distances was applied.

C. Calculating Relative and Absolute Free Energies of Hydration. Calculating the free energies of hydration was not in any way a part of the parameter fitting procedure, but rather a test of the parameters produced as discussed above. Therefore, we believe that the high quality of the results reflects the genuinely adequate underlying physical model.

The thermodynamic cycle used to calculate relative hydration energies between species A and B is shown in Figure 2.

From this cycle, the relative free energy of hydration is

$$\Delta\Delta G_{\mathrm{hyd}} = \Delta G_{\mathrm{hyd}}(\mathrm{B}) - \Delta G_{\mathrm{hyd}}(\mathrm{A}) = \Delta G(\mathrm{A} \rightarrow \mathrm{B,liq}) - \Delta G(\mathrm{A} \rightarrow \mathrm{B,gas}) \qquad (20)$$

The statistical perturbation theory was used to calculate the differences of free energies between solvated and gas-phase species A and B (ΔG(A→B, liq) and ΔG(A→B, gas), respectively). When B was set to nothing, the absolute free energy of hydration of species A was obtained. This was done to methane to anchor the other hydration free energies and to obtain their absolute values.

Table 1. Three-Body Energy Deviations from Quantum Mechanical Data, kcal/mol

molecule	deviation, kcal/mol rms	maximum	maximum energy
H_2O	0.2682	0.3688	−0.5620
CH_3OH	0.2080	0.3588	−0.3700
NMA	0.3651	0.5902	−0.6293

Table 2. Two-Body Energy Deviations from Quantum Mechanical Data, kcal/mol

molecule	deviation, kcal/mol rms	maximum	maximum energy
H_2O	1.1482	1.4819	−12.4741
CH_3OH	1.1627	1.3804	−12.1679
NMA	1.7319	2.6399	−14.9697

To calculate the ΔG(A→B, liq) and ΔG(A→B, gas), the standard statistical perturbation theory procedure was used. Differences between atoms of molecules A and B were switched on according to a parameter, $0 < \lambda < 1$, with $\lambda = 0$ corresponding to A, and $\lambda = 1$ to B. Then the interval from 0 to 1 was divided into a number of subintervals, and for each point between two intervals, a corresponding mixture of molecules A and B was created. The difference in free energies between systems corresponding to such points i and j was calculated according to eq 21:[17]

$$\Delta G(i \rightarrow j) = -RT\ln(\langle \exp[(-E_j - E_i)/RT] \rangle_i) \qquad (21)$$

Here, the brackets, $\langle ... \rangle_i$, signify averaging of the value inside the brackets over the configurational space of the mixed system at point i, and E_i and E_j are energies of the mixed systems i and j. In other words, the free energy difference between the molecular systems A and B is the thermodynamic average of their energy differences, and the whole change from A to B is broken into a number of steps in order to speed up the convergence.

The averaging was performed with Monte Carlo calculations for a single solute molecule in a water box, thus, corresponding to infinitely dilute solutions. The simulations proceeded as described in the previous subsection, except that a number of water molecules equal to the number of non-hydrogen atoms in the solute were removed (for example, for the methanol to ethane perturbation, the number of water molecules was equal to 214 instead of the pure water box of 216).

III. Results and Discussion

A. Fitting Electrostatic Part of the Force Field. As described above, fitting polarizabilities to the three-body energies and charges to the interaction energies with dipolar probes were the first two steps of the POSSIM force field production. While the further fine-tuning did lead to some adjustments, we still view reproducing the quantum mechanical three- and two-body energies as an important part of the force field validation. Listed in Tables 1 and 2 are three-body and two-body energies resulting from the final

Table 3. Dipole Moments, in Debye

molecule	dipole moment, experimental[a]	dipole moment, calculated
H_2O	1.85	1.98
CH_3OH	1.69	2.00
NMA	3.7	3.81

[a] Ref 18.

parameter sets. The following conclusions can be drawn from these data. The rms deviations of the three-body energies are within ca. 0.37 kcal/mol, with the largest deviation corresponding to the case of N-methylacetamide. This result is not surprising, given that parameter fitting for amides is by no means a straightforward business. For example, we had to build three separate parameter sets for the formamide, acetamide, and N-methylacetamide while producing the previous version of the polarizable force field (PFF0).[9,12] The acetamide parameters were ultimately used for the protein backbone.[9] Likewise, the two-body energies display a similar trend, with the maximum rms deviation (for NMA) being ca. 1.73 kcal/mol. While the absolute value of this error seems to be large, it constitutes only about 11.5% of the two-body energy itself. It is very likely that this error is, in fact, not much greater than that resulting from the imperfections of the quantum mechanical calculations as such.

As an added test, we have compared the POSSIM dipole moments with their experimental values in Table 3. It can be seen that the results are rather close and do not suffer from the overpolarization which is characteristic for the nonpolarizable fixed-charges force field. For example, the TIP4P monomer dipole moment is 2.18 D. This overpolarization is necessary to account for the increased dipole moment in polar media (such as bulk water), but it creates a nonphysical situation in the gas phase. This problem is resolved by applying the polarization formalism.

B. Alkane Parameters. It can be noticed that results for methane, ethane, propane, and butane are not present in Tables 1−3. The nonbonded parameters (including the electrostatics) for these systems were produced in a different fashion. Following the approach employed in creating the previous generation of the polarizable force field,[12] we set the electrostatic charges on alkane atoms to be the same as in the OPLS-AA. That is, each aliphatic hydrogen had a charge of 0.06 electrons, and the sp^3 carbons were charged by −0.24, −0.18, or −0.12 electron units, depending on the number of hydrogens attached to the carbon. Since the $-CH_n-$ groups are essentially spherically symmetric, the magnitudes of the charges do not make a significant difference. This has been shown previously by simulating pure liquid and hydrated saturated hydrocarbons.[19] Furthermore, the σ and ε Lennard-Jones parameters were also set to their OPLS-AA values, 3.5 Å and 0.066 kcal/mol for carbons and 2.5 Å and 0.030 kcal/mol for hydrogens, respectively. Finally, the isotropic polarizabilities for the carbon atoms in the POSSIM PFF are the same as had been previously found to perform well in the PFF0.[12] For these atoms, the inverse polarizability $\alpha^{-1} = 0.5069$ Å^{-3}. No polarizabilities were assigned to the aliphatic hydrogen atoms.

Table 4. Torsional Energy, in kcal/mol

molecule	dihedral	angle values	energy, QM[a]	energy, PFF
C_2H_6	H−C−C−H	0°	2.7762	2.7751
		60°	0.0000	0.0000
C_3H_8	H−C−C−C	0°	3.1298	3.1294
		60°	0.0000	0.0000
C_4H_{10}	C−C−C−C	0°	5.5337	5.5371
		60°	0.5751	0.5752
		120°	3.2104	3.2069
		180°	0.0000	0.0000
CH_3OH	H−C−O−H	0°	1.0643	1.0647
		60°	0.0000	0.0000
NMA	C−C−N−C	0°	2.0170	2.0185
		180°	0.0000	0.0000
	H−C−N−C	0°	0.0000	0.0000
		45°	0.0878	0.0328
		60°	0.0049	0.0480
	H−C−C−N	0°	0.0000	0.0000
		45°	0.2093	0.2154
		60°	0.2509	0.2551

[a] LMP2/cc-pVTZ(-f), this work.

Therefore, the nonbonded parameters for the alkanes were not refitted. However, the torsional coefficients for ethane, propane, and butane, H−C−C−H, H−C−C−C, and C−C−C−C, torsions were refitted to reproduce quantum mechanical energy profiles.

C. Torsional Coefficients. Torsional energy coefficients have been fitted to reproduce LMP2/cc-pVTZ(-f) energy profiles as described in the Methods section. The results of performing this task are shown in Table 4. Our original goal was to achieve a ca. 0.1 kcal/mol agreement with the quantum mechanical results, but the final accuracy was much better than that. The reason is that the molecules which we considered contained no coupled torsions (such as, for example, protein backbone ϕ and Ψ). This was true even for the NMA molecule. Therefore, we were dealing with one-dimensional torsional profiles, and adjusting the three Fourier coefficients for each torsion was enough for achieving the high level of accuracy.

D. Gas-Phase Dimerization Energies and Interatomic Distances. The next step in our force field development, and perhaps the first truly significant test of the new force field, was reproducing gas-phase binding energies and geometries of the complexes. Given in Tables 5 and 6 are the results of comparing our polarizable force field application with quantum mechanical and experimental data. As described in the previous section, the quantum mechanical results, unless otherwise noted, are taken from applying our previously developed extrapolation technique utilizing calculations with LMP2/cc-pVTZ(-f) and LMP2/cc-pVQZ basis sets.

Table 5 contains results for the aliphatic hydrocarbons (as represented by methane), water, and methanol. Unlike in generating the previous version of the polarizable force field (PFF0[12]), we have calculated energies for both the homodimer of methanol and its heterodimer with water. Moreover, both methanol dimers with water in which the latter serves as an electron donor and acceptor have been considered. This was also our routine when dealing with the NMA. We plan to follow this pattern in the future development of the POSSIM force field. Methane−water dimers were not considered, as no true hydrogen bond is formed in this case.

Table 5. Gas-Phase Dimerization Energies (kcal/mol) and Distances (Å) for H_2O, CH_4, and CH_3OH

	energy				distance			
system	QM[a]	OPLS[a]	PFF0[b]	PFF[c]	QM[a]	OPLS[a]	PFF0[b]	PFF[c]
H_2O-H_2O O···O	−5.02	−6.78	−5.54	−4.52	2.91	2.68	2.88	2.91
CH_4-CH_4 C···C	−0.5[d]	−0.48	−0.44	−0.48	3.7[d]	3.77	3.86	3.76
MeOH−MeOH O···O	−5.59	−6.41	−5.63	−5.59	2.80	2.78	2.81	2.81
MeOH−OH_2 O···O	−4.90[e]			−5.12	2.86[e]			2.80
MeHO−HOH O···O	−4.77[f]			−4.93	2.91[f]			2.90

[a] Refs 7 and 12. Results for water are from ref 20. [b] Ref 12. PFF0 data for water are from ref 21. [c] This work. [d] Ref 23. [e] Ref 22. [f] Methanol−water results: from this work, LMP2 with cc-pVTZ(-f) to cc-pVQZ extrapolation, as in ref 15.

Table 6. Gas-Phase Dimerization Energies (kcal/mol) and Distances (Å) for NMA (cis and trans conformations)

	energy				distance			
system	QM[a]	OPLS[a]	PFF0[a]	PFF[b]	QM[a]	OPLS[a]	PFF0[a]	PFF[b]
NMA(t)−NMA(t) N···O	−8.53 −7.2[c] [d]	−7.96	−5.24	−7.89	2.97	2.81	2.84	2.92
NMA(c)−NMA(c) N···O	−14.4	−11.3	−14.1	−9.48	2.92	2.83	2.90	2.92
NMA(t)−OH_2 HN···O	−5.82 −4.8[c]			−4.86	3.03			3.02
NMA(t)−HOH CO···O	−7.91 −7.0[c]			−7.10	2.91			2.79

[a] Ref 12 for NMA(c) and this work for NMA(t). [b] This work. [c] Ref 24. [d] Ref 25.

As demonstrated by the results in Table 5, the dimerization energies for all of the systems were within 0.5 kcal/mol of their quantum mechanical counterparts. For the methanol−methanol dimer, the agreement is much better than that. A water−water interaction energy lower than −5 kcal/mol has been suggested previously; therefore, the relatively high error in this case is probably not a reason for concern. The distances between the heavy atoms are within ca. 0.1 Å of the quantum mechanical results. Overall, the developed PFF performs well in reproducing these gas-phase dimerization properties.

It can also be seen that the POSSIM PFF is showing a slight improvement over the PFF0. And the improvement with comparison to the fixed-charges OPLS-AA is quite dramatic. For example, as shown in Table 5, the water−water dimerization energy is overestimated by almost 2 kcal/mol if the OPLS-AA model is used. This is natural, as fixed-charges models need to overpolarize individual molecules in order to capture the overall polarization increase in the bulk polar media (such as, for example, bulk water). Therefore, our fast second-order polarization technique does not require any sacrifice of computational accuracy as compared to the full-scale polarizable model (PFF0) and outperforms the fixed-charges one.

Given in Table 6 are results of simulating NMA dimers in the gas phase. Fitting parameters for the NMA molecule is generally a more complex task. One reason for that is that quantum mechanical data for such dimerization energies are often less reliable. Shown in the table are just two values of the *trans*-NMA dimerization energy. And the full range of this piece of data which can be found in the literature is actually considerably wider. Perhaps this is why some authors choose to simply ignore the NMA gas-phase dimerization energy when producing or reparameterizing a force field.[25] We took a middle-road approach in regard to this matter. While we did use the quantum mechanical value for the NMA(t)−NMA(t) dimer, we did not pose as strong restrictions on the PFF results as we did for the other systems. We considered an error of 0.64 kcal/mol to be acceptable, given the uncertainty of the quantum mechanical results themselves. At the same time, we adhered to the stronger criteria when reproducing geometries of the complexes, with the error being within ca. 0.1−0.15 Å range.

The new POSSIM PFF performs better in reproducing the NMA(t) and NMA− water dimers than both force fields used for comparison, the fixed-charges OPLS-AA and the previously generated PFF0. While the OPLS-AA NMA(t) dimer has an energy which is slightly closer to the QM value than the POSSIM PFF, this is achieved at the cost of underestimating the N···O distance by about 0.16 Å versus the 0.05 Å error in the new PFF. As far as the PFF0 is concerned, it provides much better values for both the energy and the distance for the NMA(c) dimer but significantly underestimates both the distance and the magnitude of the energy for the dimer of the NMA in its trans form. The reason is that we considered only the *cis*-NMA dimer when the PFF0 was developed, since its dimerization energy is stronger than for the NMA(t). At the same time, the *trans*-NMA is much less abundant than the cis form, and it is the NMA(t) which forms the building block of protein and peptide backbones. This is why we choose to use the NMA(t) dimer as the fitting target in this work, and our results definitely show an improvement over the PFF0 in this respect.

E. Simulations of Pure Liquids. Calculated heats of vaporization and molecular density are shown in Table 7. It follows from the values of the average errors that the overall performance of the POSSIM PFF is superior to both that of the OPLS-AA and PFF0. The average PFF deviation of the heats of vaporization is 0.083 kcal/mol versus the 0.0.129 kcal/mol for the OPLS-AA and 0.240 kcal/mol for the PFF0. The average error in the molecular volume is 1.485 $Å^3$, which is a noticeable improvement, as compared to the PFF0, and is slightly better than that for the OPLS-AA. Moreover, the worst (although still

Table 7. Liquid State Heats of Vaporization (kcal/mol) and Molecular Volumes (Å^3), at 25°C, Except for CH_4 (−161.49 °C), C_2H_6 (−88.63 °C), C_3H_8 (−42.1 °C), C_4H_{10} (−0.5 °C), and NMA (100 °C)

	ΔH_{vap}				V			
system	exptl[a]	OPLS[a]	PFF0[b]	PFF[c]	exptl[a]	OPLS[a]	PFF0[b]	PFF[c]
H_2O	10.51	10.46	10.54	10.577 ± 0.055	30.0	30.0	30.15	29.891 ± 0.226
CH_4	1.96	2.19	1.89	2.209 ± 0.026	62.8	57.2	62.2	56.863 ±0.381
C_2H_6	3.62	3.44	3.32	3.456 ± 0.063	91.5	92.5	94.4	92.140 ± 0.571
C_3H_8	4.49	4.55	4.79	4.499 ± 0.129	126.0	125.2	123.9	126.096 ± 0.903
C_4H_{10}	5.35	5.43	5.62	5.335 ± 0.164	160.3	161.3	157.2	163.433 ± 1.416
NMA	13.3	13.6	13.9	13.319 ± 0.372	135.9	133.9	128.3	135.670 ± 0.566
CH_3OH	8.95	8.95	8.84	9.005 ± 0.140	67.7	68.3	67.0	67.447 ± 0.535
average error		0.129	0.240	0.083		1.574	2.450	1.485

[a] Refs 7 and 12. [b] Refs 12 and 21. [c] This work.

Table 8. Free Energies of Hydration, kcal/mol

process	ΔG_{hyd}, experiment	ΔG_{hyd}, PFF (this work)
nothing → CH_4	2.00[a]	2.115 ± 0.057
CH_4 → C_2H_6	−0.17[a]	−0.317 ± 0.040
C_2H_6 → CH_3OH	−6.94[b]	−6.748 ± 0.053
absolute ΔG_{hyd} (from above data)		
CH_4	2.00[a]	2.115 ± 0.057
C_2H_6	1.83[a]	1.798 ± 0.070
CH_3OH	−5.11[b]	−4.950 ± 0.088

[a] Ref 19. [b] Ref 26.

acceptable) results for the new PFF are obtained for methane and ethane, which are probably not the most relevant pure liquids considered, given their very low boiling temperatures of −161.49 and −88.63 °C, respectively. At the same time, the density of the NMA was improved very significantly as compared to the PFF0, and the NMA molecule is extremely important as the building block for protein backbones. Overall, all three models perform well in simulating the pure liquids involved. Therefore, we have clearly shown that our second-order approximation for calculating induced electrostatic dipoles is completely adequate, as is it capable of reproducing both gas-phase and liquid properties at a good level of accuracy.

F. Free Energy Perturbations. Presented in this subsection are the results of calculating relative and absolute free energies of hydration for methane, ethane, and methanol. In contrast with the work described above, this part of the project was carried out as a test and not a parametrization; thus, we did not refit any parameters to obtain good results. Therefore, these calculations can serve as a test of the physical robustness of the model. The successful results presented here have a double value of validating both the force field parameters and the overall methodological and POSSIM software robustness in calculating free energy changes with the statistical perturbation theory.

The calculated values of the free energies are shown in Table 8. We carried out three sets of these calculations: transforming methanol into ethane (by converting the −OH group into a methyl), shrinking the methyl group into an aliphatic hydrogen to mutate ethane into methane, and finally, making the methane molecule disappear altogether by converting all of the atoms into dummy ones. This last step permits us to anchor all of the results into the energy of solvation of a dummy atom (i.e. zero); thus, we can obtain the absolute hydration energies for all of the species involved.

According to the data in Table 8, the agreement of our results with those of the experiment is excellent. The hydration energy of methane is correct within 0.115 kcal/mol. The nonobvious trend of the slight hydration energy drop between methane and ethane is reproduced correctly, even though it is exaggerated by about 0.15 kcal/mol (which actually leads to the absolute hydration energy of ethane being in a better agreement with the experiment). The difference in hydration energies between methanol and ethane are correct within ca. 0.21 kcal/mol. Overall, the absolute values of these energies are within 0.13 kcal/mol of their experimental counterparts. It is important that this agreement was achieved without any specific fitting. These results represent a test of the parameters produced as described above; therefore, they attest to the robustness of the physical model and the POSSIM software.

IV. Conclusions

We have utilized a previously suggested fast polarization formalism to create a software package named POSSIM and produced polarizable force field parameters for several molecular systems relevant in organic and biological simulations. The agreement with gas-phase and liquid experimental data has been found to be very good. Moreover, we tested POSSIM by calculating absolute free energies of hydration of methane, ethane, and methanol and determined them to be within 0.13 kcal/mol of their experimental counterparts. We believe that this verifies the robustness of the parameters and the software. We view this work as the first step in creating an extensive set of force field parameters to be used in organic and biophysical simulations, including modeling of proteins and protein−ligand complexes. Furthermore, we have been able to successfully use a procedure for utilizing both quantum mechanical and experimental data as fitting targets. We hope our results will further prove the importance of treating the electrostatic polarization explicitly, when building empirical force fields, and will contribute to making polarizable calculations more affordable computationally.

Acknowledgment. The project described was supported by Grant Number R01GM074624 from the National Institute of General Medical Sciences. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of

General Medical Sciences or the national Institutes of Health. The authors express gratitude to Schrödinger, LLC for the Jaguar and PFF software.

Supporting Information Available: POSSIM force field parameters. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) See, for example: (a) Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 4177–4178. (b) Jiao, D.; Zhang, J. J.; Duke, R. E.; Li, G. H.; Schneiders, M. J.; Ren, P. Y. *J. Comput. Chem.* **2009**, *30*, 1701–1711.

(2) (a) MacDermaid, C. M.; Kaminski, G. A. *J. Phys. Chem. B* **2007**, *111*, 9036–9044. (b) Click, T. H.; Kaminski, G. A. *J. Phys. Chem. B* **2009**, *113*, 7844–7850.

(3) Veluraja, K.; Margulis, C. J. *J. Biomol. Struct. Dyn.* **2005**, *23*, 101–111.

(4) Ji, C.; Mei, Y.; Zhang, J. Z. H. *Biophys. J.* **2008**, *95*, 1080–1088.

(5) Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A.; Cao, Y. X.; Murphy, R. B.; Zhou, R.; Halgren, T. *J. Comput. Chem.* **2002**, *23*, 1515–1531.

(6) Rick, S. W.; Stuart, S. J.; Berne, B. J. *J. Chem. Phys.* **1994**, *101*, 6141–6156.

(7) Kaminski, G. A.; Zhou, R.; Friesner, R. A. *J. Comput. Chem.* **2003**, *24*, 267–276.

(8) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.

(9) Maple, J. R.; Cao, Y. X.; Damm, W.; Halgren, T. A.; Kaminski, G. A.; Zhang, L. Y.; Friesner, R. A. *J. Chem. Theory Comput.* **2005**, *1*, 694–715.

(10) (a) Straatsma, T. P.; McCammon, J. A. *Chem. Phys. Lett.* **1990**, *167*, 252–254. (b) Straatsma, T. P.; McCammon, J. A. *Chem. Phys. Lett.* **1991**, *177*, 433–440. (c) Roux, B. *Chem. Phys. Lett.* **1993**, *212*, 231–240.

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

(12) Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A. *J. Phys. Chem. A* **2004**, *108*, 621–627.

(13) (a) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098–3100. (b) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789.

(14) (a) *Jaguar*, v3.5; Schrödinger, Inc.: Portland, OR, 1998. (b) *Jaguar*, v4.2; Schrödinger, Inc.: Portland, OR, 2000.

(15) Kaminski, G. A.; Maple, J. R.; Murphy, R. B.; Braden, D; Friesner, R. A. *J. Chem. Theory Comput.* **2005**, *1*, 248–254.

(16) Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007–1023.

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

(18) Caldwell, J. W.; Kollman, P. A. *J. Phys. Chem.* **1995**, *99*, 6208–6219.

(19) Kaminski, G. A.; Duffy, E. M.; Matsui, T.; Jergensen, W. L. *J. Phys. Chem.* **1994**, *98*, 13077–13082.

(20) Mattsson, A. E.; Mattsson, T. R. *J. Chem Theory Comput.* **2009**, *5*, 887–894.

(21) Kaminski, G. A. *J. Phys. Chem. B* **2005**, *109*, 5884–5890.

(22) Tsuzuki, S.; Uchimaru, T.; Matsumura, K.; Mikami, M.; Tanabe, K. *J. Chem. Phys.* **1999**, *110*, 11906–11910.

(23) Takatani, T.; Sherrill, C. D. *Phys. Chem. Chem. Phys.* **2007**, *9*, 6106–6114.

(24) Dixon, D. A.; Dobbs, K. D.; Valentini, J. J. *J. Phys. Chem.* **1994**, *98*, 13435–13439.

(25) Xie, W.; Pu, J.; MacKerell, A. D.; Gao, J. *J. Chem. Theory Comput.* **2007**, *3*, 1878–1889.

(26) Gallicchio, E.; Zhang, L. Y.; Levy, R. M. *J. Comput. Chem.* **2002**, *23*, 517–529.

CT900409P