

Fluctuating Charge, Polarizable Dipole, and Combined Models: Parameterization from *ab Initio* Quantum Chemistry

Harry A. Stern,[†] George A. Kaminski,[†] Jay L. Banks,[†] Ruhong Zhou,[‡] B. J. Berne,^{*,†} and Richard A. Friesner[†]

Department of Chemistry and Center for Biomolecular Simulation, Columbia University, New York, New York 10027, and Schrödinger, Inc., 411 Hackensack Avenue, 10th floor, Hackensack, NJ 07601

Received: November 20, 1998; In Final Form: March 15, 1999

We performed studies of fluctuating charge, fluctuating dipole, and combined models for substituted benzenes and concluded that dipoles are necessary to avoid errors in important cases. Force fields incorporating fluctuating dipoles for alanine, serine, and phenylalanine were developed that accurately reproduce both relative conformational energies and cooperative many-body energies as given by *ab initio* quantum chemical calculations. The polarization model was fit to quantum chemical calculations of changes in the electrostatic potential (ESP) induced by applied perturbations. The electrostatic model was completed by adding fixed charges fit to the zero-field quantum chemical ESP. All intramolecular and Lennard–Jones parameters, and some torsional parameters, were taken from the OPLS-AA force field of Jorgensen and co-workers. Key torsional parameters were refit to quantum chemical structures and energies.

I. Introduction

The inclusion of polarizability in molecular modeling force fields is a crucial step necessary for building a genuinely accurate and predictive computational tool capable of reliably representing complex systems. In a previous paper,¹ we presented our first attempt to extend polarizable models from their typical small-molecule applications to a relatively large organic molecule, alanine dipeptide. This work employed the fluctuating charge (FQ) model of Berne and co-workers² to represent the polarizability of the molecule and the OPLS-AA force field of Jorgensen and co-workers³ for the remainder of the force field (with some minor modifications). A novel approach to fitting quantum chemical response data was introduced and shown to provide, for the most part, excellent predictions of many-body energies. At the same time, the new force field yielded respectable results in reproducing the relative energies of 10 conformations of alanine tetrapeptide as compared with accurate quantum chemistry. These results provided encouraging evidence that the construction of a polarizable force field directly from quantum chemistry was a feasible goal.

However, as was noted in ref 2, the polarization response of a point-charge-only model is limited in certain cases; e.g., for a planar molecule, there is no out-of-plane response. Reference 1 described a number of important cases in which such a model appears to be qualitatively inadequate, for example, bifurcated hydrogen bonds to oxygen. If one is going to go to the trouble of implementing a detailed microscopic description of polarization, with its attendant complexities and additional computational cost, it is imperative that the description of the polarization achieve a minimum standard of accuracy. As will be shown below, a point-charge model fails by this criterion. Motivated by these observations, we have extended our approach to include inducible dipoles as well as fluctuating point charges. Representing molecular polarizability by a system of induced dipoles

on atoms is a well-established technique^{9–15} and has been generalized to all orders of multipoles by Stone.¹⁶

It is possible to apply different levels of description to different atoms, or to different regions of large molecules. Such an approach will probably be necessary for efficient calculations on large systems. As computational cost is roughly proportional to the square of the number of fluctuating charge sites, but 9 times the square of the number of dipole sites, it is clear that a model based primarily on fluctuating charges is preferable. We have performed a comparison for small-molecule model systems that indicates that the best description for a given cost is a combined FQ–dipole model that can describe both interatomic charge transfer and atomic polarization. For the peptide models presented in this paper, however, we are interested in the ability of a fitting protocol based on quantum chemistry to accurately predict many-body responses and relative conformational energetics, without regard to computational cost, and have thus chosen a simpler representation consisting of dipoles only. Work on combined models for water, simple organic liquids, and peptides is continuing in our laboratory.

Once the model for polarization has been constructed and validated, several steps remain to assemble a molecular mechanics force field, a task we have carried out for the amino acids alanine, serine, and phenylalanine. The electrostatic model is completed by fitting to zero-field density functional theory (DFT) calculations using a large basis set. Stretch, bend, and van der Waals parameters are taken from the OPLS-AA force field. The final step is refitting of the torsional parameters; we have developed a method for fitting coupled torsions, based on gradient weighting, which appears to provide reliable results in an automated fashion. The completed force fields are then evaluated by their ability to reproduce accurate quantum chemical conformational data; accurate and robust results are obtained.

The present work, while encouraging, does not yet represent a definitive demonstration that a robust force field, which will provide accurate numbers in both the gas phase and the

[†] Columbia University.

[‡] Schrödinger, Inc.

condensed phase, has been produced. That will require comparing condensed-phase simulations with experiment (e.g., calculation of liquid-state densities and heats of vaporization) and, simultaneously, reproducing quantum chemical binding energies for molecular pairs. Work investigating these issues is currently in progress; while the preliminary results appear to be reasonable, it is as yet too early to draw serious conclusions.

The paper is organized as follows. Section II presents the formalism for a general linear response model containing both fluctuating charges and dipoles and describes our computational approach to fitting the parameters to quantum chemical data. Section III presents results for different polarization models in reproducing quantum chemical three-body energies for a wide range of cases. Section IV discusses assembly of the remainder of the force field, and section V gives results obtained for alanine, serine, and phenylalanine. Section VI, the conclusion, discusses future directions.

II. Linear Response Model for Polarization

A. Functional Form. Consider a polarizable system represented by fluctuating charges q_A on a set of atoms A and induced dipoles $\bar{\mu}_B$ on a (possibly overlapping or identical) set of atoms B. The system is also subject to an “external” electrostatic potential $\phi^0(\vec{r})$ with gradient $-\vec{E}^0(\vec{r})$. The superscript zero denotes that this electrostatic potential and field do not arise from the fluctuating charges or dipoles, but from some other source, for instance, a set of fixed charges.

Each fluctuating charge q_A has a self-energy $\chi_A q_A + 1/2 J_A q_A^2$, where χ_A and J_A are parameters corresponding to the atomic electronegativity and hardness.^{2,4} The interaction with the external potential gives a term $\phi_A^0 q_A$ where ϕ_A^0 is the value of the external potential at site A. Pairs of fluctuating charges $q_A, q_{A'}$ give rise to an interaction energy $q_A J_{AA'} q_{A'}$ where $J_{AA'}$ depends on the distance between sites A and A'. For instance, if we assume the interaction is Coulombic, then

$$J_{AA'} = \frac{1}{|\vec{r}_{AA'}|} \quad (1)$$

where $\vec{r}_{AA'} = \vec{r}_A - \vec{r}_{A'}$, is the displacement vector from site A' to site A.

The dipolar terms are quite similar. If α_B is the polarizability tensor for atom B, then an induced dipole $\bar{\mu}_B$ has a self-energy $1/2 \bar{\mu}_B \cdot \alpha_B^{-1} \bar{\mu}_B$.¹² In addition, $\bar{\mu}_B$ interacts with the external field giving a term $-\vec{E}_B^0 \cdot \bar{\mu}_B$, where \vec{E}_B^0 is the value of the field at site B. Pairs of dipoles $\bar{\mu}_B, \bar{\mu}_{B'}$, give rise to an interaction energy $\bar{\mu}_B \cdot \mathcal{J}_{BB'} \bar{\mu}_{B'}$, where $\mathcal{J}_{BB'}$ depends on the locations of sites B and B' and must be a dyadic so that the interaction energy is independent of the choice of coordinate system. If we assume the interaction is Coulombic, then

$$\mathcal{J}_{BB'} = \frac{1}{|\vec{r}_{BB'}|^3} \left(1 - 3 \frac{\vec{r}_{BB'} \vec{r}_{BB'}}{|\vec{r}_{BB'}|^2} \right) \quad (2)$$

Finally, the fluctuating charges and dipoles interact (if they are on different sites). Each pair of fluctuating charges $q_A, \bar{\mu}_B$ gives an interaction energy $q_A \bar{J}_{AB} \bar{\mu}_B$. As before \bar{J}_{AB} depends on the locations of sites A and B and in this case is a vector. Assuming the interaction is Coulombic,

$$\bar{J}_{AB} = \frac{\vec{r}_{AB}}{|\vec{r}_{AB}|^3} \quad (3)$$

The total electrostatic energy due to the fluctuating charges and dipoles may therefore be written

$$U = \sum_A (\chi_A + \phi_A^0) q_A - \sum_B \vec{E}_B^0 \cdot \bar{\mu}_B + \frac{1}{2} \sum_A J_A q_A^2 + \frac{1}{2} \sum_B \bar{\mu}_B \cdot \alpha_B^{-1} \bar{\mu}_B + \frac{1}{2} \sum_{A \neq A'} q_A J_{AA'} q_{A'} + \frac{1}{2} \sum_{B \neq B'} \bar{\mu}_B \cdot \mathcal{J}_{BB'} \bar{\mu}_{B'} + \sum_{AB} q_A \bar{J}_{AB} \bar{\mu}_B \quad (4)$$

It is convenient to define $J_{AA} \equiv J_A$ and $\mathcal{J}_{BB} \equiv \alpha_B^{-1}$; in this case the energy may be written slightly more simply,

$$U = \sum_A (\chi_A + \phi_A^0) q_A - \sum_B \vec{E}_B^0 \cdot \bar{\mu}_B + \frac{1}{2} \sum_{AA'} q_A J_{AA'} q_{A'} + \frac{1}{2} \sum_{BB'} \bar{\mu}_B \cdot \mathcal{J}_{BB'} \bar{\mu}_{B'} + \sum_{AB} q_A \bar{J}_{AB} \bar{\mu}_B \quad (5)$$

Let us now define $N_A + 3N_B$ dimensional vectors \mathbf{q} and \mathbf{v} and an $N_A + 3N_B$ by $N_A + 3N_B$ matrix \mathbf{J} , where N_A is the number of fluctuating charges and N_B is the number of dipoles,

$$\begin{aligned} \mathbf{q} &\equiv (q_A, \bar{\mu}_B) \\ \mathbf{v} &\equiv (\chi_A + \phi_A^0, -\vec{E}_B^0) \\ \mathbf{J} &\equiv \begin{pmatrix} J_{AA'} & \bar{J}_{AB'} \\ \bar{J}_{A'B}^\dagger & \mathcal{J}_{BB'} \end{pmatrix} \end{aligned} \quad (6)$$

Then, eq 5 may be written succinctly as a matrix equation,

$$U = \mathbf{v}^\dagger \mathbf{q} + \frac{1}{2} \mathbf{q}^\dagger \mathbf{J} \mathbf{q} \quad (7)$$

For any given set of atomic electronegativities χ_A and values for the external potential and field ϕ^0 and \vec{E}^0 at the sites A and B, the fluctuating charges and induced dipoles are determined by minimizing eq 5 with respect to each variable $q_A, \bar{\mu}_B$. It can be seen that, in the case of an all-dipole system, this is equivalent to imposing the usual self-consistent field requirement on the induced dipoles. If, as in this case, there are no constraints on the variables, then minimizing leads to a set of linear equations whose solution is

$$\mathbf{q} = -\mathbf{J}^{-1} \mathbf{v} \quad (8)$$

Constraints on the fluctuating charges, such as the requirement that each molecule remain neutral, may be handled by the method of Lagrange multipliers, or by a transformation to a reduced set of unconstrained coordinates \mathbf{q}' , where $\mathbf{C}^\dagger \mathbf{q}' = \mathbf{q}$ for some matrix \mathbf{C} . In this case, the solution is given by

$$\mathbf{q} = -\mathbf{C}^\dagger (\mathbf{C} \mathbf{J} \mathbf{C}^\dagger)^{-1} \mathbf{C} \mathbf{v} \quad (9)$$

We note that the response $\Delta \mathbf{q}$ to any additional perturbation $\Delta \mathbf{v}$, for instance, an external, applied electrostatic potential or field from additional charges, is simply

$$\Delta \mathbf{q} = -\mathbf{J}^{-1} \Delta \mathbf{v}$$

$$\Delta \mathbf{q} = -\mathbf{C}^\dagger (\mathbf{C} \mathbf{J} \mathbf{C}^\dagger)^{-1} \mathbf{C} \Delta \mathbf{v} \quad (10)$$

for unconstrained and constrained coordinates, respectively. The response to external perturbations does not depend on \mathbf{v} , that

is, on the electronegativities and original fixed charges we have placed in the system.

A polarization model for a given molecule therefore involves a specification for the elements of the matrix \mathbf{J} , that is, the interactions between pairs of fluctuating charges and dipoles. We note that if we are interested simply in how well an electrostatic model of charges and dipoles at various sites can reproduce quantum chemical responses for a single, fixed conformation of a molecule, then we could specify the elements of the inverse matrix \mathbf{J}^{-1} directly. In this paper, we have followed this latter approach for the fluctuating charge, polarizable dipole, and combined FQ–dipole models for substituted benzenes. The transferable polarization models we fit for the peptides consist only of dipoles, so we need only specify the coupling between pairs of dipoles.

Several approaches to determining these interactions exist in the literature. Applequist⁹ treated all couplings Coulombically (without screening), which may be problematic, especially for flexible models. Most of the polarizable dipole models proposed for water and other simple liquids^{21,22} have treated intermolecular interactions Coulombically and omitted close-range intramolecular interactions, although Levy and co-workers¹⁵ and, recently, Burnham et al.²⁰ have followed the approach of Thole¹⁰ and introduced a screening function based upon spatially smearing out the point multipoles. (The fluctuating charge models have followed a similar approach, determining close-range interactions by Coulomb or overlap integrals over Gaussian^{5,8} or Slater^{2,4} distributions.) Most of these approaches rely on empirical molecular polarizabilities or adjust the parameters in order to reproduce condensed-phase properties.

Our approach is similar; we have treated long-range interactions Coulombically and short-range interactions by a scaled or “screened” Coulomb interaction. This screened interaction takes the simple form of a Coulomb interaction at some fixed, effective distance $K_{BB'}$ rather than the actual distance $r_{BB'}$, and we will use it whenever $r_{BB'}$ is less than $K_{BB'}$:

$$\mathcal{J}_{BB'} = \begin{cases} \frac{1}{r_{BB'}^3} \left(1 - 3 \frac{\vec{r}_{BB'} \cdot \vec{r}_{BB'}}{|\vec{r}_{BB'}|^2} \right), & r_{BB'} > K_{BB'} \\ \frac{1}{K_{BB'}^3} \left(1 - 3 \frac{\vec{r}_{BB'} \cdot \vec{r}_{BB'}}{|\vec{r}_{BB'}|^2} \right), & r_{BB'} < K_{BB'} \end{cases} \quad (11)$$

In practice, a cutoff radius was fit for each atom type having a dipole, and the effective distance $K_{BB'}$ was given by the sum of the radii for the two atom types involved. The diagonal elements \mathcal{J}_{BB} are treated as parameters; as before, these correspond to the inverse of the polarizability tensor α_B for dipoles. For computational convenience, we have restricted these polarizabilities to be isotropic.

B. Parameterization. A standard way to obtain atomic charges for molecular mechanics simulations is to perform a least-squares fit to the electrostatic potential (ESP), calculated by quantum chemical techniques, at a set of gridpoints around the molecule.²³ We extend this idea to fitting parameters for polarizable models in the following way. For each conformation of a molecule, we apply a series of external electrostatic perturbations $\Delta v(\vec{r})$ to the molecule. As earlier, we will associate a vector $\Delta \mathbf{v}$ with each perturbation, whose elements $\{\Delta \phi_A, -\Delta E_B\}$ are the value of the external potential at sites A, at which there are fluctuating charges, and its gradient at sites B, at which there are polarizable dipoles. For each perturbation, we compute the change in the ESP at a set of gridpoints, as given by ab initio calculations on the perturbed and unperturbed conforma-

tions and choose parameters of the linear response model so as to best fit these deviations.

Let \vec{r}_k be the location of gridpoint k , and let $\Delta \Phi$ be a vector whose k th element, $\Delta \Phi_k$, is the change, due to the perturbation, of the ESP given by the model at gridpoint k . For a combined model with fluctuating point charges q_A and induced dipoles $\vec{\mu}_B$,

$$\Delta \Phi_k = \sum_A \frac{1}{|\vec{r}_{kA}|} \Delta q_A + \sum_B \frac{\vec{r}_{kB}}{|\vec{r}_{kB}|^3} \cdot \Delta \vec{\mu}_B \quad (12)$$

In matrix notation, $\Delta \Phi = \mathbf{R} \Delta \mathbf{q}$, with $\Delta \mathbf{q} = \{\Delta q_A, \Delta \vec{\mu}_B\}$ as before, and with the matrix elements R_{ki} of \mathbf{R} given by $R_{kA} = 1/|\vec{r}_{kA}|$ and $R_{kB} = \vec{r}_{kB}/|\vec{r}_{kB}|^3$. If we let \mathbf{w} be a diagonal matrix whose k th element is the square root of the weight we wish to assign in the fit to gridpoint k , and $\Delta \Phi_{\text{ab initio}}$ be a vector whose k th element is the change in the ab initio ESP at gridpoint k , then we may express the fitting problem as the minimization of a cost function χ^2 :

$$\chi^2 = \sum ||\mathbf{w} \Delta \Phi - \mathbf{w} \Delta \Phi_{\text{ab initio}}||^2 \quad (13)$$

$$= \sum ||\mathbf{w} \mathbf{R} \Delta \mathbf{q} - \mathbf{w} \Delta \Phi_{\text{ab initio}}||^2 \quad (14)$$

$$= \sum ||\mathbf{w} \mathbf{R} \mathbf{J}^{-1} \Delta \mathbf{v} + \mathbf{w} \Delta \Phi_{\text{ab initio}}||^2 \quad (15)$$

where the sum is over all applied perturbations. Again, we assume unconstrained variables \mathbf{q} ; if there are constraints, we may handle them as described earlier. An important requirement for the fitting is that the matrix \mathbf{J} must be positive definite, so that eq 8 specifies a minimum and a “polarization catastrophe”¹⁰ does not occur.

It is known that instabilities may arise in ESP fitting if charges are poorly determined by the set of gridpoints; for instance, in the case of charges on “buried” atoms far inside the van der Waals surface.²⁴ Instabilities might show up in poor values for χ^2 , overly large induced charges or dipole moments, or small or negative eigenvalues in the matrix \mathbf{J} . As in previous work,¹ we address this problem by zeroing poorly-determined modes via singular value decomposition, a standard technique.²⁵ We note that other approaches are possible, for instance, introducing penalty functions to constrain some of the charges to target values.²⁴ Let $\mathbf{U} \mathbf{s} \mathbf{V}^\dagger$ be the singular value decomposition of the ESP-fitting matrix $\mathbf{w} \mathbf{R}$, and σ be a diagonal matrix with either a one or a zero for the i th diagonal element, depending on whether or not we wish to zero the i th mode. The cost function then becomes

$$\chi^2 = \sum ||\mathbf{U} \mathbf{s} \mathbf{V}^\dagger \mathbf{J}^{-1} \Delta \mathbf{v} + \mathbf{U} \sigma \mathbf{U}^\dagger \mathbf{w} \Delta \Phi_{\text{ab initio}}||^2 \quad (16)$$

$$= \sum ||\mathbf{s} \mathbf{V}^\dagger \mathbf{J}^{-1} \Delta \mathbf{v} + \sigma \mathbf{U}^\dagger \mathbf{w} \Delta \Phi_{\text{ab initio}}||^2 \quad (17)$$

since the matrix \mathbf{U} is column-orthonormal. Since χ^2 is a nonlinear function of the parameters we intend to fit (for the all-dipole peptide models, these are the diagonal elements and cutoff distances, as described above), we will need to use a nonlinear optimization method. It is helpful to make an approximation to χ^2 that is linear in the matrix elements of \mathbf{J} so as to obtain an initial guess. This we do by multiplying both terms inside the vector norm in eq 17 by $\mathbf{J} \mathbf{V} \mathbf{s}^{-1}$, obtaining

$$\begin{aligned} \chi_{\text{approx}}^2 &= \sum ||\Delta \mathbf{v} + \mathbf{J} \mathbf{V} \mathbf{s}^{-1} \sigma \mathbf{U}^\dagger \mathbf{w} \Delta \Phi_{\text{ab initio}}||^2 \\ &= \sum ||\Delta \mathbf{v} + \mathbf{J} \Delta \mathbf{q}_{\text{ab initio}}||^2 \end{aligned} \quad (18)$$

Here, $\Delta\mathbf{q}_{\text{ab initio}}$ is a vector containing those values of the variables \mathbf{q} that best fit the changes in the ab initio ESP (with the specified modes removed). For example, if we had a model consisting only of point charges, $\Delta\mathbf{q}_{\text{ab initio}}$ would consist of the differences between the usual ESP-fit charges for the perturbed and unperturbed molecules.

The approximation given by eq 18 is exact if \mathbf{J} is orthogonal, which, of course, will not be true in general. In any case χ^2_{approx} may be minimized by simple linear least-squares techniques with respect to the parameter contributions to \mathbf{J} , yielding an initial guess for them. Subsequently, we minimize eq 17 directly, first by simulated annealing and then by conjugate gradient local minimization. For larger molecules with many parameters, it was necessary in some cases to constrain diagonal elements to be greater than zero and off-diagonal elements to be smaller than some upper bound in order to avoid excessive wandering in parameter space. In general, the quality of the fit as given by χ^2 was not overly sensitive to many details of the fitting process, for instance, the exact value of the parameter bounds, the annealing schedule, or even whether or not the first initial guess was performed.

When we fit a complex molecule with many parameters and constraints on the variables \mathbf{q} , we have found it necessary to enforce the constraints during the fitting process by a transformation to a smaller set of unconstrained coordinates, rather than by the method of Lagrange multipliers, to avoid problems discussed in ref 1. This issue did not arise for the peptide models of the current work, since they were all-dipole models with no constraints placed on the variables.

An alternate method of fitting may be applied in the case that we do not want to obtain a set of transferable parameters, and do not constrain long-range interactions to be Coulombic, but are interested simply in how well the functional form specified by eq 5 can reproduce quantum chemical results for a single, fixed conformation of a molecule. In this case, we fit elements of \mathbf{J}^{-1} , rather than contributions to \mathbf{J} , by minimizing the cost function

$$\chi^2 = \sum \|\Delta\mathbf{q} - \Delta\mathbf{q}_{\text{ab initio}}\|^2 \quad (19)$$

$$= \sum \|\mathbf{J}^{-1}\Delta\mathbf{v} + \Delta\mathbf{q}_{\text{ab initio}}\|^2 \quad (20)$$

which may be done by simple linear least-squares techniques. This second method of fitting, to which we refer hereafter as the ‘‘inverse fit’’, is primarily useful for comparing the performance of different physical models, since it does not yield transferable parameters.¹

III. Results for Polarization Responses

Using the above methodology, we fit polarization models to the following molecules: a number of substituted benzenes, alanine dipeptide, serine dipeptide, and phenylalanine dipeptide. For the substituted benzenes, we used the inverse fit to parameterize three different models for the purpose of comparing them: a model with fluctuating charges on all atoms, designated FQ; a model with dipoles on all atoms, designated PD; a model with fluctuating charges on all atoms and dipoles on heavy atoms, designated FQ+PD. For alanine dipeptide, we used a polarizable-dipole model with dipoles on all carbons and nitrogens and fit the diagonal parameters $\mathcal{J}_{\text{BB}} = \alpha_{\text{B}}^{-1}$, constraining these to be isotropic, as well as effective distances K_{BB} , which were taken to be a sum of cutoff distances for atom

TABLE 1: Electrostatic Parameters for Polarizable-Dipole Models of Alanine, Serine, and Phenylalanine Dipeptides.

atom type	fixed charge (e)	polarizability (\AA^3)	cutoff radius (\AA)
backbone			
amide hydrogen	0.253		
hydrogen on α -carbon	0.027		
carbonyl carbon	0.037	1.524	1.278
α -carbon	0.077	0.763	1.130
amide nitrogen	-0.356	0.764	0.707
carbonyl oxygen	-0.496	0.369	0.744
terminal acetyl hydrogen	-0.038		
terminal <i>N</i> -methylamide hydrogen	0.098		
terminal <i>N</i> -methylamide carbon	-0.114	2.037	2.940
terminal acetyl carbonyl carbon	0.410	1.253	2.003
terminal acetyl methyl carbon	0.172	1.991	2.701
alanine side chain			
hydrogen on β -carbon	0.024		
β -carbon	0.008	2.069	2.787
serine side chain			
hydrogen on β -carbon	0.102	0.450	3.000
β -carbon	-0.065	0.669	0.928
oxygen	-0.381	0.405	0.011
hydroxyl hydrogen	0.324	0.232	1.618
phenylalanine side chain			
hydrogen on β -carbon	0.046		
β -carbon	0.012	0.861	0.000
γ -carbon	-0.028	1.567	2.508
other aromatic carbons	-0.103	0.705	0.000
aromatic hydrogens	0.104		

types B and B'. We transferred the backbone parameters thus obtained to serine and phenylalanine dipeptides. For these dipeptides, we also placed dipoles on all heavy side-chain atoms (as well as side-chain hydrogens in the case of serine) and fit additional parameters. The electrostatic parameters for the dipeptides are given in Table 1.

All calculations of the ab initio ESP deviations were performed at the Hartree-Fock level (6-31G** basis set), using the Jaguar electronic structure code.²⁶ Geometries were taken from conformations minimized at the HF/6-31G** level: six conformations for alanine dipeptide, seven for serine dipeptide, and three for phenylalanine dipeptide. For each conformation, we generated between 20 and 30 perturbing fields by placing dipolar probes at hydrogen-bonding positions and other locations around the molecule at a distance of 1.5 \AA or greater. The probes consisted of oppositely charged point charges of magnitude 0.7815 e spaced 0.5774 \AA apart.

To test the polarization models we calculated three-body energies, $U(3) = U_{123} - U_{12} - U_{13} - U_{23} + U_1 + U_2 + U_3$, for trimers consisting of the molecule and two dipolar probes placed at hydrogen-bonding positions around the molecule. These three-body energies depend only on the matrix \mathbf{J} and are independent of \mathbf{v} .

We compared the model predictions with ab initio calculations, again at the HF/6-31G** level. Results for the fit and three-body energies are given in Table 2 and Figures 1–3. We note a systematic improvement in the accuracy of the three-body energies in going from an FQ-only model to a PD-only model to a combined model.

The largest errors in the point-charge-only models occur when the probes form bifurcated hydrogen bonds, or are located above and below a ring (see Figure 4). The three-body energy of a point-charge-only model is systematically too low in these cases, as shown in Figures 1 and 5. The same effect has been observed in bifurcated hydrogen bonds formed by fluctuating-point-charge models of water.⁷ A possible explanation is that this systematic error results from estimating a nearly isotropic response with a response directed primarily along bonds, such as the bond

TABLE 2: Response Data for a Series of Molecules^a

molecule	fit RMS (kcal/(mol e))	three-body RMS (kcal/mol)	three-body max (kcal/mol)
C ₆ H ₄ NH ₂ NO ₂ (FQ)	0.593	0.520	1.606
C ₆ H ₄ OHNH ₂ (FQ)	0.682	0.811	2.794
C ₆ H ₄ OHNO ₂ (FQ)	0.651	0.924	3.115
C ₆ H ₅ NH ₂ (FQ)	0.672	0.728	1.723
C ₆ H ₅ NO ₂ (FQ)	0.701	0.536	1.301
C ₆ H ₅ OH (FQ)	0.667	1.078	2.971
C ₆ H ₄ NH ₂ NO ₂ (PD)	0.084	0.378	1.043
C ₆ H ₄ OHNH ₂ (PD)	0.080	0.233	0.952
C ₆ H ₄ OHNO ₂ (PD)	0.098	0.320	1.049
C ₆ H ₅ NH ₂ (PD)	0.083	0.454	1.039
C ₆ H ₅ NO ₂ (PD)	0.101	0.436	0.907
C ₆ H ₅ OH (PD)	0.082	0.283	0.780
C ₆ H ₄ NH ₂ NO ₂ (FQ + PD)	0.120	0.244	0.665
C ₆ H ₄ OHNH ₂ (FQ + PD)	0.111	0.195	0.697
C ₆ H ₄ OHNO ₂ (FQ + PD)	0.142	0.220	0.641
C ₆ H ₅ NH ₂ (FQ + PD)	0.123	0.296	0.692
C ₆ H ₅ NO ₂ (FQ + PD)	0.136	0.278	0.602
C ₆ H ₅ OH (FQ + PD)	0.144	0.248	0.698
alanine dipeptide (PD)	0.123	0.167	0.324
serine dipeptide (PD)	0.182	0.161	0.501
phenylalanine dipeptide (PD)	0.273	0.134	0.302

^a "Fit RMS" is the weighted root mean square deviation between $\Delta\Phi$ and $\Delta\Phi_{\text{ab initio}}$, at all gridpoints, for all perturbations. "Three-body RMS" and "three-body max" are the root-mean-square deviation and maximum deviation between ab initio three-body energies and model three-body energies, for trimers consisting of the molecule and two dipolar probes placed at hydrogen-bonding positions.

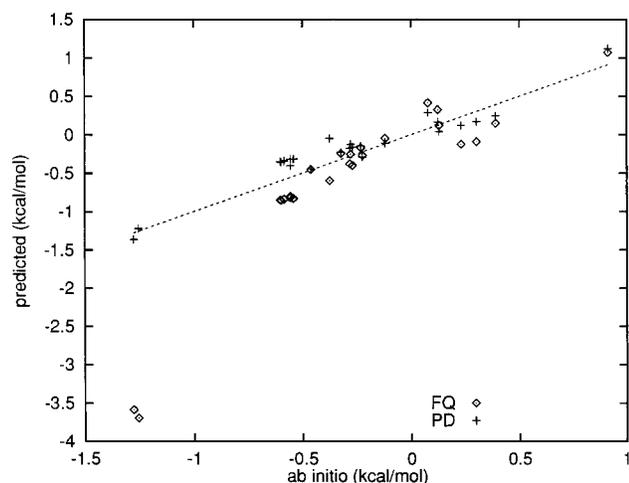


Figure 1. Comparison between three-body energies for trimers consisting of alanine dipeptide and two dipole probes. The two lowest energy trimers are bifurcated hydrogen bonds. Two models are compared: fluctuating charges only (FQ) and polarizable dipoles only (PD).

between the carbon and oxygen of a carbonyl group involved in a bifurcated hydrogen bond or the bonds between the carbons and hydrogens of an aromatic ring. Consider a very simple model for the configurations shown in Figure 4, a polarizable center with isotropic polarizability α , representing the molecule, surrounded by two fixed point charges q , representing the probes. Each charge is located at a distance r , and the charges and polarizable center make an angle θ . The three-body energy of this system is

$$U_{\text{isotropic}} = -\frac{q^2\alpha}{r^4} \cos \theta \quad (21)$$

If we now restrict the center to polarizing only along the line bifurcating the charges, while keeping the average polarizability

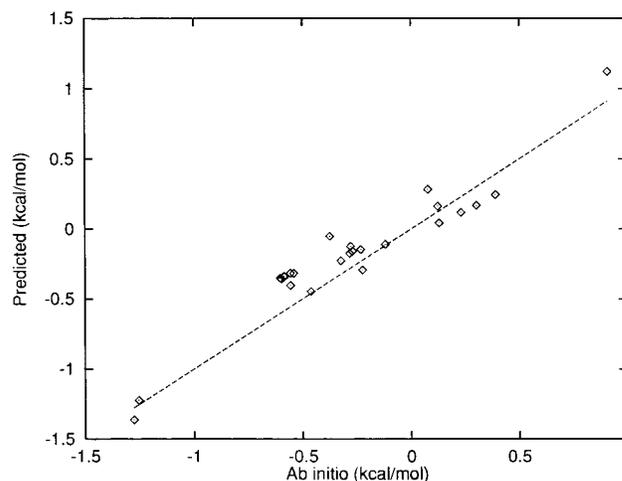


Figure 2. Three-body energies for trimers consisting of serine dipeptide and two dipole probes.

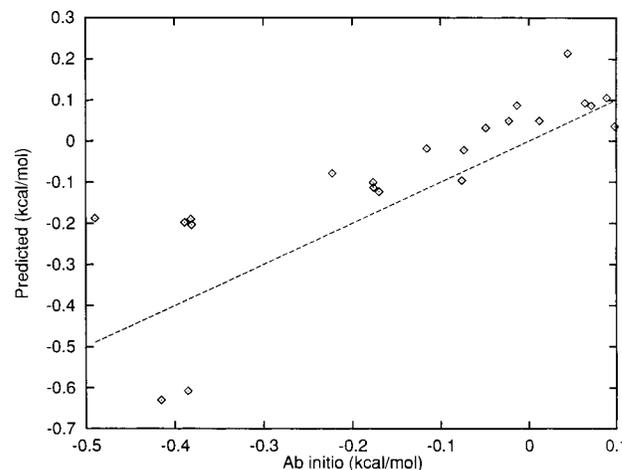


Figure 3. Three-body energies for trimers consisting of phenylalanine dipeptide and two dipole probes.

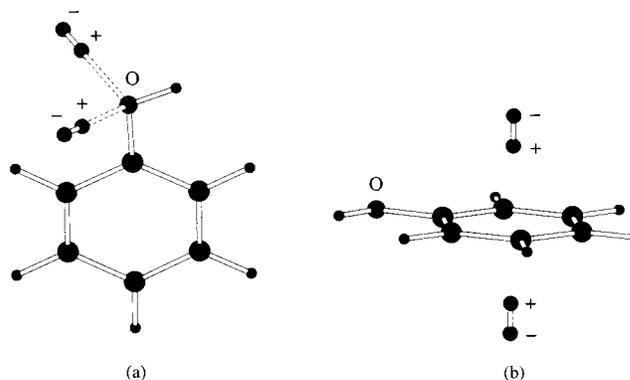


Figure 4. Two cases in which a point-charge-only model fails: (a) a bifurcated hydrogen bond and (b) probes above and below an aromatic ring.

Tr $\alpha/3$ constant, the three-body energy of the system becomes

$$U_{\text{restricted}} = -\frac{3q^2\alpha}{r^4} \left(\frac{1 + \cos \theta}{2} \right) \quad (22)$$

which is always less than $U_{\text{isotropic}}$. Assuming an aromatic ring is better described by a central polarizability ($\theta \approx 180^\circ$) than by a set of polarizabilities around its edges ($\theta \approx 90^\circ$), eq 21 also provides an explanation for why the three-body energy is

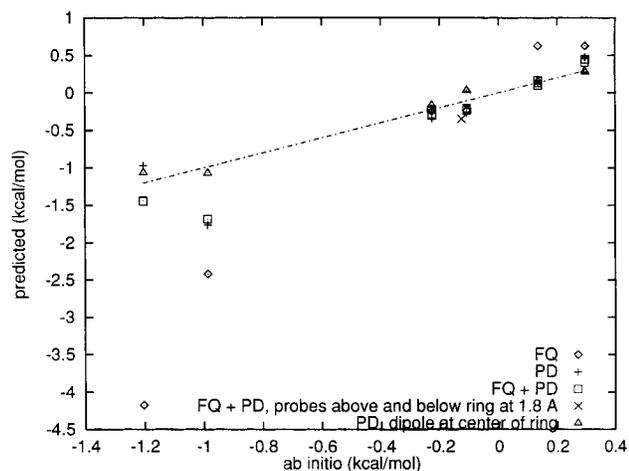


Figure 5. Three-body energies for trimers made up of C_6H_5OH and two dipolar probes placed at hydrogen-bonding positions at a distance of 1.5 Å and greater. The lowest energy trimer is a bifurcated hydrogen bond, while the second lowest corresponds to probes placed directly above and below the ring. Various models are compared: fluctuating charges only (FQ), polarizable dipoles only (DP), a combined model (FQ + DP), and a polarizable dipole model with an additional site at the center of the ring.

still too small for a model with dipoles situated only on atoms. This suggests placing an additional dipole at the center of the ring, which indeed fixes the problem entirely (see Figure 5).

Though off-atom sites are a complication, and we have not included them in the peptide models yet, they appear necessary in order to faithfully reproduce this test case, although it is unclear what degree of accuracy is needed: it might be argued that the test at hand is unreasonably demanding in that approaches to aromatic rings are rarely, if ever, as close as 1.5 Å and the error is less significant at longer distances, e.g., 1.8 Å.

IV. Assembling the Force Field

After the polarization was fit, we completed the electrostatic models for alanine, serine, and phenylalanine dipeptides by adding fixed point charges on all atoms. The interaction of the fixed charges with the polarizabilities was taken to be purely Coulombic if they were separated by at least two bonds and zero otherwise. The values of the fixed charges were determined by a least-squares fit of the total ESP of the model, from both the fixed charges and the resulting polarization response, to the ESP as given by zero-field DFT/B3LYP calculations^{27,28} with the cc-pVTZ(-f) basis set, again using the Jaguar program.²⁶ For alanine and phenylalanine dipeptides, the ESP from the lowest energy conformation was fit. For serine dipeptide, slightly better results were obtained by fitting to the ESP from the second-lowest energy conformation. Future work will incorporate fitting zero-field ESPs from several conformations simultaneously. Fixed charges are given along with the polarization parameters in Table 1.

Applications such as gas-phase energy minimizations, conformational search, binding calculations, and liquid-phase simulations require defining a full force field, including, in our case, bond stretches, angle bends, torsions, and van der Waals interactions. While concentrating on building the electrostatic model, we intended to use existing OPLS-AA parameters for the remainder of the force field whenever possible. We chose this force field because it performs very well for a wide variety of bioorganic systems. Although our electrostatic model has a different functional form, it should produce similar electrostatic

TABLE 3: New Torsional Parameters (kcal/mol)

angle	V_1	V_2	V_3
backbone ϕ/ψ			
C(O)–N–C–C(O)	–1.381	–0.044	–2.138
N–C–C(O)–N	1.615	2.495	–1.439
C(O)–N–C–C	–0.474	–0.060	2.203
C–C–C(O)–N	2.206	0.436	1.209
serine χ_1/χ_2			
N–C–C–O	0.127	0.692	1.214
C(O)–C–C–O	–0.650	–0.912	0.000
C–C–O–H	0.796	–1.421	1.035

interactions on average if it is to have the same accuracy in reproducing bulk properties of liquids.

As we have stated above, we adopted all stretch, bend, and Lennard-Jones parameters from OPLS-AA, with the exception of key torsional parameters. The torsional energy is given by

$$E_{\text{torsion}} = \sum_i \frac{1}{2} [V_1(1 + \cos \phi_i) + V_2(1 - \cos 2\phi_i) + V_3(1 + \cos 3\phi_i)] \quad (23)$$

where the sum is over all the dihedral angles ϕ_i , and V_1 , V_2 , and V_3 are the Fourier coefficients. Although we adopted the OPLS-AA values of the coefficients for some of the torsions, we treated several key cases differently. These cases were the coupled ϕ and ψ angles in the peptide backbone and the serine dipeptide χ_1 and χ_2 side-chain angles. The torsions were treated differently because in the OPLS-AA approach electrostatic 1–4 interactions are produced by permanent point charges and scaled by 0.5, while in our methodology, the same interactions (a) are not scaled at all and (b) are truly nonadditive and emerge as a result of taking into account the total electrostatic field produced by the permanent charges and induced dipoles together. It should be noted that such torsions may be affected by long-range (longer than 1–4) electrostatic interactions which, even if similar to the OPLS-AA case on average, are expected to behave differently locally. Finally, we used OPLS-AA torsional parameters for the other dihedrals, for example, the methyl group rotations, because most of the methyl hydrogens lack polarizability in our present model and their permanent electrostatic charges are rather small in magnitude.

We will describe the torsional fitting procedure we used for the backbone parameters and for the serine side-chain rotations in a separate paper,²⁹ so we outline it briefly here. First of all, potential energy minima were identified for the energy surface defined by the coupled torsions and ab initio geometry optimizations were carried out for the minima. Then, for each of the minima, a series of restrained energy minimizations was done. The coupled angles were constrained so that one of them was the same as at the energy minimum and the other assumed different values. Thus, one-dimensional slices of the potential energy surface were obtained for each of the conformers. The same minimization series were then carried out with the force field, and the data from the former and the latter were used in least-square fitting of the torsional parameters for the angles involved. The points on the energy surface employed in the fitting were weighted according to estimated gradient values at the ab initio surface, with the weights decreasing exponentially with increase of the gradient magnitudes. Results presented in the next section illustrate the advantages of this procedure. All the ab initio energies were obtained at the LMP2/cc-pVTZ(-f)//HF/6-31G** level of accuracy. New torsional parameters are given in Table 3.

TABLE 4: Alanine Dipeptide Conformational Energies (kcal/mol)

	ab initio ^a	OPLS	this work	this work
ϕ/ψ fitting		no ^b	no ^b	yes
C7eq	0.00	-0.31	0.16	0.00
C5	0.95	1.01	0.55	0.78
C7ax	2.67	2.24	2.10	2.48
b2	2.75			
aL	4.31			
ap	5.51	6.19	6.32	5.88
RMS error ^c		0.43	0.54	0.22

^a LMP2/cc-pVTZ(-f)//HF/6-31G**. ^b Standard OPLS-AA torsional parameters employed. ^c For every method, position of the energy minima were uniformly shifted to achieve the lowest possible RMS deviation from the ab initio data.

TABLE 5: Alanine Tetrapeptide Conformational Energies (kcal/mol)

	ab initio ^a	OPLS	this work	this work
ϕ/ψ fitting ^b		no ^c	no ^c	yes
1 ^d	2.71	2.56	2.19	2.88
2	2.84	2.20	1.61	1.84
3	0.00	-1.57	-0.35	0.22
4	4.13	3.33	3.13	3.69
5	3.88	4.32	3.14	3.70
6	2.20	2.94	2.64	1.45
7	5.77	3.85	5.64	5.48
8	4.16	6.79	6.22	5.38
9	6.92	5.82	5.54	6.74
10	6.99	9.35	9.86	8.21
RMS error ^e		1.47	1.34	0.71

^a LMP2/cc-pVTZ(-f)//HF/6-31G**. ^b No refitting done for the tetrapeptide, dipeptide-fitted ϕ/ψ torsional parameters used. ^c Standard OPLS-AA torsional parameters employed. Parameters transferred directly from the dipeptide. ^d The minima numbering as in ref 31. ^e For every method, position of the energy minima were uniformly shifted to achieve the lowest possible RMS deviation from the ab initio data.

V. Applications: Alanine, Serine, and Phenylalanine

Reproducing relative ab initio conformational energies for alanine dipeptide was the first target of the force field assembled as described above. Energy minimizations were performed with the BOSS version 3.6 program,³⁰ modified to employ the polarizable electrostatic model. First, backbone ϕ and ψ torsional parameters were fit as described in the previous section. It should be pointed out that the fitting was done on alanine dipeptide only, and the parameters thus obtained were then used for both the dipeptide and tetrapeptide geometry minimizations, as well as in serine and phenylalanine calculations. In this way, transferability of the parameters was tested. Tables 4 and 5 show results of geometry optimizations for alanine dipeptide and tetrapeptide, respectively. Several observations can be made. First, the proposed polarizable force field performs comparably to OPLS-AA, even without the backbone torsional refitting. Second, fitting ϕ and ψ torsional parameters decreases the RMS deviations for the dipeptide by about a factor of two. Finally, torsional parameters fitted on the dipeptide potential surface perform very well in the alanine tetrapeptide energy minimizations, where the RMS energy error also dropped by nearly 50%. To summarize, the presented force field is superior to any other empirical force field in reproducing the alanine tetrapeptide conformational energies, as illustrated by Table 6.

Table 7 presents results of energy minimizations for serine dipeptide. We used the same backbone torsional coefficients as in the alanine case described above. In addition, the side chain χ_1 and χ_2 torsions were fitted to ab initio data. The procedure was exactly the same as described in the previous

TABLE 6: RMS Deviations from the LMP2/cc-pVTZ(-f)//HF/6-31G ab initio Values for Alanine Tetrapeptide Conformational Energies Minimizations^a**

force field	energy RMS (kcal/mol)
LMP2/cc-pVTZ(-f)	
this work	0.71
OPLS-FQ (ref 1)	0.94
HF/6-31G**	1.10
MMFF93	1.20
OPLS-AA(2,2)	1.47
MM3*	1.53
GROMOS	1.60
HF/cc-pVTZ(-f)	1.69
Null Hypothesis ^b	2.07
AMBER*	2.39
CHARMM 22	2.56

^a Data from this work and refs 1 and 31. ^b All the conformers are assumed to have the same energy.

TABLE 7: Serine Dipeptide Conformational Energies (kcal/mol)

conformer	ab initio ^a	OPLS-AA	this work
1	0.00	0.45	-0.05
2	3.04	3.38	3.04
3	3.99	2.97	3.76
4	4.25	4.18	4.39
5	5.56	5.60	5.60
6	7.50	7.77	7.61
RMS error ^b		0.49	0.12
V_{\max}^c (kcal/mol)		6.28	1.42

^a LMP2/cc-pVTZ(-f)//HF/6-31G**. ^b For every method, position of the energy minima were uniformly shifted to achieve the lowest possible RMS deviation from the ab initio data. ^c The greatest magnitude among the side chain torsional coefficients.

section for the ϕ/ψ coupled fitting; we used one-dimensional slices of the potential energy surface defined by the χ_1 and χ_2 angles, taking the slices at the minima.

It is clear that the new model gives a significant improvement over standard OPLS-AA results for serine. The minima now have the correct order, which was not the case for the OPLS-AA conformations 2 and 3. A smaller shift is now needed to achieve the lowest energy RMS. The RMS itself is 4 times lower than that for OPLS-AA. Finally, the greatest Fourier coefficient magnitude needed for the side-chain torsions is only 1.421 kcal/mol in the new model, versus the unusually high 6.280 kcal/mol used in OPLS-AA. This demonstrates that the potential energy surface produced by the polarizable model in hand is significantly closer to the ab initio one, even before the discrepancy is narrowed by means of the torsional energy fitting.

Table 8 displays results of geometry minimizations of phenylalanine conformers. The backbone torsional parameters were taken from the alanine dipeptide fitting. No ab initio torsional fitting was done for the side chain; instead, we set all the Fourier coefficients to zero for the C-C-C-X torsions in the χ_1 dihedral. Here again, the polarizable force field produces a potential energy surface in better agreement with ab initio calculations than standard OPLS-AA so that no torsional adjustment of this surface is needed for this angle. The results of geometry optimizations for the phenylalanine dipeptide presented in Table 8 are in excellent agreement with the ab initio data.

To summarize, the accuracy of the new models in reproducing the relative conformational energies of alanine, serine, and phenylalanine dipeptides and alanine tetrapeptide is superior to that of standard OPLS-AA, and it appears that the ab initio potential energy surface is better reproduced after including

TABLE 8: Phenylalanine Dipeptide Conformational Energies (kcal/mol)

conformer	ab initio ^a	OPLS-AA	this work
1	0.00	-0.08	0.13
2	0.65	0.49	0.68
3	1.89	2.11	1.73
RMS error ^b		0.16	0.12

^a LMP2/cc-pVTZ(-f)//6-31G**. ^b For every method, position of the energy minima were uniformly shifted to achieve the lowest possible RMS deviation from the ab initio data.

electrostatic polarization, even without refitting side-chain torsional parameters.

VI. Conclusions

We have presented a molecular mechanics force field that treats electrostatic polarization effects explicitly in the framework of a linear response model. The polarization parameters were fit to reproduce HF/6-31G** calculations of changes in the electrostatic potential resulting from perturbing fields. Fixed charges were added so as to best reproduce zero-field DFT/B3LYP/cc-pVTZ(-f) calculations of the electrostatic potential. Lennard-Jones, bond stretching, and angle bending parameters were taken directly from the OPLS-AA force field, as were some of the torsional Fourier coefficients, while the key backbone and side-chain torsional parameters were refit to ab initio energetic and structural results.

The many-body response of the peptide models is quite accurate, as was demonstrated by good agreement with ab initio calculations of the changes in the electrostatic potential due to perturbations, and with three-body energies of trimers mimicking cooperative hydrogen bonds. Studies on substituted benzenes indicated a shortcoming in point-charge-only models, systematic underestimation of many-body energies from bifurcated hydrogen bonds or charged groups above and below an aromatic ring, which is probably due to estimating an isotropic response with a response directed along a line. The error is corrected by including dipole polarizabilities. For the particular case of charges 1.5 Å directly above and below a ring, off-atom sites are required; but it is unclear whether this case is of sufficient importance to warrant the additional complication.

We have demonstrated that the proposed models are capable of reproducing alanine, serine, and phenylalanine dipeptide and alanine tetrapeptide conformational energies in good agreement with the LMP2/cc-pVTZ(-f)//HF/6-31G** ab initio results and that they are superior to OPLS-AA for those cases.

Although these results are encouraging, further work needs to be done in order to ascertain that our approach will yield robust models for condensed-phase simulations. The most

important directions to be pursued immediately are tests on a wider variety of protein residues and tests of the performance of the proposed technique in reproducing thermodynamic liquid-phase properties, e.g., heats of vaporization, densities, and free energies of solvation.

Acknowledgment. This work was funded by the National Institutes of Health under Grants GM52018 (R.A.F.), GM19961 (G.A.K.), and GM43340 (B.J.B.). H.A.S. was supported in part by an NIH training grant in molecular biophysics (5T32 GM08281). We thank Dr. Mike Beachy for ab initio data and Dr. Dean Philipp for assistance with the Jaguar program.

References and Notes

- (1) Banks, J. L.; Kaminski, G. A.; Zhou, R.; Mainz, D. T.; Berne, B. J.; Friesner, R. A. *J. Chem. Phys.* **1999**, *110*, 741.
- (2) Rick, S. W.; Stuart, S. J.; Berne, B. J. *J. Chem. Phys.* **1994**, *101*, 6141.
- (3) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225.
- (4) Rappé, A. K.; Goddard, W. A. *J. Phys. Chem.* **1991**, *95*, 3358.
- (5) Sprik, M.; Klein, M. L. *J. Chem. Phys.* **1988**, *89*, 7556.
- (6) Zhu, S.-B.; Singh, S.; Robinson, G. W. *J. Chem. Phys.* **1991**, *95*, 2791.
- (7) Liu, Y. P.; Kim, K.; Berne, B. J.; Friesner, R. A.; Rick, S. W. *J. Chem. Phys.* **1998**, *108*, 4739.
- (8) York, D.; Yang, W. *J. Chem. Phys.* **1996**, *104*, 159.
- (9) Applequist, J.; Carl, J. R.; Fung, K.-K. *J. Am. Chem. Soc.* **1972**, *94*, 2952.
- (10) Thole, B. T. *Chem. Phys.* **1981**, *59*, 341.
- (11) Rullmann, J. A. C.; van Duijnen, P. Th. *Mol. Phys.* **1988**, *63*, 451.
- (12) Ahlström, P.; Wallqvist, A.; Engström, S.; Jönsson, B. *Mol. Phys.* **1989**, *68*, 563.
- (13) Cieplak, P.; Kollman, P. A. *J. Chem. Phys.* **1990**, *92*, 6755.
- (14) Dinur, U. *J. Chem. Phys.* **1993**, *97*, 7894.
- (15) Bernardo, D. N.; Ding, Y.; Krogh-Jespersen, K.; Levy, R. M. *J. Phys. Chem.* **1994**, *98*, 4180.
- (16) Stone, A. J. *Mol. Phys.* **1985**, *56*, 1065.
- (17) Miller, K. J. *J. Am. Chem. Soc.* **1990**, *112*, 8533.
- (18) Voisin, C.; Cartier, A. *THEOCHEM (J. Mol. Struct.)* **1993**, *105*, 35.
- (19) Itskowitz, P.; Berkowitz, M. L. *J. Phys. Chem. A* **1997**, *101*, 5687.
- (20) Burnham, C. J.; Jichen, Li.; Xantheas, S. S.; Leslie, M. *J. Chem. Phys.* **1999**, *110*, 4566.
- (21) Gao, J.; Habibollahzadeh, D.; Shao, L. *J. Phys. Chem.* **1995**, *99*, 16460.
- (22) Caldwell, J. W.; Kollman, P. A. *J. Phys. Chem.* **1995**, *99*, 6208.
- (23) Momany, F. *J. Chem. Phys.* **1978**, *82*, 592.
- (24) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269.
- (25) See, for example: Press et al. *Numerical Recipes in FORTRAN: The Art of Scientific Computing*, 2nd ed.; Cambridge University Press: New York, 1992.
- (26) *Jaguar v3.0*; Schrödinger, Inc.: Portland, OR, 1997.
- (27) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098.
- (28) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.
- (29) Kaminski, G. A. In preparation.
- (30) Jorgensen, W. L. *BOSS*, Version 3.6; Yale University: New Haven, CT, 1995.
- (31) Beachy, M. D.; Chasman, D.; Murphy, R. B.; Halgren, T. A.; Friesner, R. A. *J. Am. Chem. Soc.* **1997**, *119*, 5909.