

## Efficient Simulation Method for Polarizable Protein Force Fields: Application to the Simulation of BPTI in Liquid Water

Edward Harder, Byungchan Kim, Richard A. Friesner, and B. J. Berne\*

*Department of Chemistry and Center for Bimolecular Simulation, Columbia University,  
3000 Broadway, New York, New York 10027*

Received October 18, 2004

**Abstract:** A methodology for large scale molecular dynamics simulation of a solvated polarizable protein, using a combination of permanent and inducible point dipoles with fluctuating and fixed charges, is discussed and applied to the simulation of water solvated bovine pancreatic trypsin inhibitor (BPTI). The electrostatic forces are evaluated using a generalized form of the P3M Ewald method which includes point dipoles in addition to point charge sites. The electrostatic configuration is propagated along with the nuclei during the course of the simulation using an extended Lagrangian formalism. For the system size studied, 20000 atoms, this method gives only a marginal computational overhead relative to nonpolarizable potential models (1.23–1.45) per time step of simulation. The models employ a newly developed polarizable dipole force field for the protein<sup>1</sup> with two commonly used water models TIP4P-FQ and RPOL. Performed at constant energy and constant volume (NVE) using the velocity Verlet algorithm, the simulations show excellent energy conservation and run stably for their 2 ns duration. To characterize the accuracy of the solvation models the protein structure is analyzed. The simulated structures remain within 1 Å of the experimental crystal structure for the duration of the simulation in line with the nonpolarizable OPLS-AA model.

### I. Introduction

The goal in force field development for biomolecular systems is to retain chemical accuracy while taking advantage of computational expediency by employing the simplest potential function. Expressing the electrostatic potential energy using a system of fixed point charges interacting via Coulomb's law is certainly simple and is the approach taken for the most popular models used in biomolecular force field simulation.<sup>2–4</sup> However, such nonpolarizable force fields do not reflect the dependence of a molecule's electronic structure on its environment. This dependency is clearly manifest in water where the magnitude of the average dipole moment is approximately 40% larger in the liquid compared to the gas phase. For homogeneous systems, such as neat fluids, the exclusion of polarization to model the electrostatic energy

may be sufficient for some purposes. However, the electrostatic environments found in solvated biomolecules range from nonpolar near hydrophobic residues to highly polar in the vicinity of hydrophilic and charged residues to a nearly bulk water like environment far from the protein. A rescaling of the partial charges to reflect the mean field response is one way to deal with the average effects of condensed phase environments; however, inhomogeneous systems with spatially varying fields necessitate the explicit inclusion of polarization to properly treat the electrostatic potential.

How best to incorporate polarization in a simple manner is an ongoing quest. The distributed polarizability analysis of Stone<sup>5</sup> is an approach that incorporates highly distributed inducible sites occupied by high order point multipoles. The main drawback to such an approach is the additional complexity of the potential function and the corresponding increase in computational cost. Recently the particle mesh Ewald method has been extended to include multipole

\* Corresponding author phone: (212)854-2186; fax: (212)854-7454; e-mail: berne@chem.columbia.edu.

interactions up to hexadecapole-hexadecapole.<sup>6</sup> Although providing a considerable improvement in efficiency relative to regular multipole Ewald,<sup>7</sup> the algorithm is still a factor of 8.5 slower than simply using point charges alone.

The most common approach to including polarization in a simple force field describes the electrostatic configuration using a system of fixed point charges/dipoles and inducible dipoles.<sup>8–18</sup> Another approach, referred to as fluctuating charges (FQ), uses variable charge magnitudes to model polarization and has been growing in popularity due to the ease of its implementation and related computational speed.<sup>19–24</sup> Recent successful efforts have led to parametrized models for small molecules that use a combination of fixed charges/dipoles, fluctuating charges, and inducible dipoles using techniques for deriving electrostatic parameters from ab initio electronic structure calculations.<sup>25,26</sup> Using this QM technique, Kaminski et al.<sup>1</sup> have now developed a complete polarizable dipole model for proteins that has shown good accuracy in gas-phase experiments.

Before the Kaminski model can be used as a predictive tool, it will be necessary to validate and refine the model for simulation in liquid water. To do so it is critical to develop an efficient formalism for simulating polarizable condensed phase biomolecular systems. It is the long range electrostatic interactions in biomolecular systems which make these simulations computationally intensive. To minimize surface effects we use periodic boundary conditions. Spherical or minimum image truncation of the long-range electrostatic forces is a method that reduces the computational cost of the simulation but gives rise to unphysical effects.<sup>27,28</sup> The Ewald sum<sup>29</sup> provides a tractable solution to the accurate evaluation of the electrostatic forces but has a computational complexity of  $O(N^{3/2})$ . For system sizes between  $10^4 < N < 10^5$ , that are necessary to simulate solvated proteins, mesh based Ewald methods such as SPME<sup>30</sup> and P3M Ewald<sup>31–33</sup> have been shown to be particularly effective. These algorithms have favorable  $O(N \log N)$  scaling and lead to approximately an order of magnitude improvement in the computational cost of evaluating the electrostatic forces for systems the size of 20 000 atoms (the relative merits and similarities of the SPME vs P3M Ewald methodology have been discussed at length elsewhere<sup>32,34</sup>). Further computational gains may be realized by using multiple time scale integration algorithms that allow for the evaluation of the expensive long range electrostatic interactions less frequently than in standard Verlet integration schemes.<sup>33,35</sup> The result of these advances is a formalism for atom-detail nonpolarizable potential functions that allows for the MD simulation of solvated proteins on nanosecond time scales using reasonable computational resources.

Explicit inclusion of polarizability in a simple potential function adds additional computational complexity which must be solved in order to realize size scales that are currently accessible to nonpolarizable molecular models. The additional computational burden is 3-fold: (1) Use of inducible point dipole sites in addition to fixed or variable charges requires additional charge-dipole and dipole-dipole interactions. (2) Resolving the electrostatic configuration and therefore the field at each molecular dynamics time step

necessitates the self-consistent solution to a system of coupled linear equations. (3) A transparent application of a multiple time scale algorithm in a fashion similar to that applied to nonpolarizable potential models is not straightforward. Recent developments have laid the groundwork for efficient large scale simulations of polarizable systems. In this article we focus on addressing the first two points in constructing an efficient simulation methodology. Work on the problem of combining multiple time scale integration schemes with polarizable potential functions is ongoing in our lab.

In a following study<sup>66</sup> we will be interested in studying the dynamic properties of water solvent in the vicinity of the protein. To ensure an accurate evaluation of these properties our simulations are conducted at constant energy and constant volume free of artificial perturbations necessary to simulate in the isothermal/isobaric ensemble.<sup>57,58</sup>

The article is organized as follows. In section 2.1 we introduce the polarizable models based on the inducible dipole model of Kaminski et al.<sup>1</sup> for the protein. We choose two solvent models for comparison, a fluctuating charge water model, TIP4P-FQ,<sup>19</sup> and an inducible dipole water model, RPOL.<sup>8</sup> To efficiently evaluate the electrostatic potential for a system of charges and dipoles we have generalized the P3M Ewald method in a fashion similar to the methodology developed by Toukmaji et al.<sup>36</sup> To efficiently resolve the electrostatic configuration and therefore the nuclear forces at each molecular dynamics time step, an extended Lagrangian method<sup>19,37</sup> combined with the generalized P3M Ewald method is used to dynamically propagate the electrostatic variables during the course of the simulation. This is discussed in sections 2.2.1 and 2.2.2. In section 3 we apply this technology to the simulation of bovine pancreatic trypsin inhibitor in water. The computational complexity and simulation accuracy is discussed in sections 3.2.1 and 3.2.2. As an initial study of the quality of the polarizable protein model and the proposed solvation models, the structure of the protein is compared with the experimental structures in section 3.2.3.

## II. Methodology

**A. Model.** The approach taken to include polarization in the force fields applied in this study replaces the usual fixed point charge representation for the electrostatic energy with a combination of fixed and variable point charges and dipoles that respond to perturbations in the electric field according to a parametrized potential energy. The charges and dipoles are located relative to the atomic positions of the molecules, either coincident with the atom position or on off-atom virtual sites.

Adding inducible point dipoles to a system of fixed charges is the most common method for introducing explicit polarizability into a molecular force field. The energy for an induced dipole moment on site  $i$  is

$$E_{\mu_i} = \gamma_i \cdot \mu_i + \frac{1}{2} \mu_i \cdot \alpha_i^{-1} \cdot \mu_i \quad (1)$$

where  $\gamma_i$  and the component of the polarizability tensor of the dipole site,  $\alpha_i$ , are treated as fitting parameters. The

parameter  $\gamma_i$  is a means of introducing a permanent dipole moment on the isolated site  $i$ . Through a simple transformation eq 1 can be expressed as the familiar self-energy of the induced dipole relative to the isolated site.

The fluctuating charge model<sup>19</sup> introduces variable charges that respond to fluctuations in the electrostatic potential according to the principle of electronegativity equalization. By this principle the charges will distribute so that the electronegativity on each variable charge site is the same subject to appropriate charge constraints.<sup>19</sup> The energy of creating a fluctuating charge is

$$E_{q_i} = \chi_i q_i + \frac{1}{2} q_i^2 J_i^0 \quad (2)$$

where the Mulliken electronegativity,  $\chi_i$  and  $J_i^0$  are treated as fitting parameters. The parameter  $J_i^0$  is twice the hardness of the electronegativity of the isolated site.<sup>19</sup>

The electrostatic potential energy in an interacting system relative to a system of isolated molecules can be expressed as

$$V^{el} = E_{elec} + E_\mu + E_q - E_{gp} \quad (3)$$

where  $E_{gp}$  is the system energy in the gas phase and  $E_{elec}$  is the electrostatic energy resulting from the interaction of different sites:

$$\begin{aligned} E_{elec} &= \sum_i \sum_{j \neq i} \left[ \frac{1}{2} q_i J(r_{ij}) q_j + q_i \mathbf{S}_{ij} \cdot \boldsymbol{\mu}_j + \frac{1}{2} \boldsymbol{\mu}_i \cdot \mathbf{T}_{ij} \cdot \boldsymbol{\mu}_j \right] \\ E_\mu &= \sum_i E_{\mu i} \\ E_q &= \sum_i E_{q i} \end{aligned} \quad (4)$$

If the electrostatic sites are well separated, the coupling terms can be expressed as Coulomb interactions,  $J(r_{ij}) = 1/(r_{ij})$ ,  $\mathbf{S}_{ij} = \mathbf{r}_{ij}/(r_{ij}^3)$ , and the tensor  $\mathbf{T}_{ij} = \mathbf{1}/(r_{ij}^3) - 3\mathbf{r}_{ij}\mathbf{r}_{ij}/(r_{ij}^5)$ . At short distances the point multipole approximation for the electron charge distribution breaks down, and the above Coulombic potential diverges. Where deemed necessary screening functions are used at close intermolecular site distances.<sup>26</sup> Intramolecular electrostatic interaction between neighbors (1,2 interactions) and one neighbor removed (1,3) are omitted in the model for the polarizable protein. For water, the (1,2) and (1,3) distances are fixed, and the Coulombic interactions are treated as either additional fitted electrostatic parameters<sup>19</sup> or omitted.<sup>8</sup>

The equilibrium charge/dipole configuration is determined at each set of nuclear coordinates by minimizing the potential energy with respect to the electrostatic variables subject to a charge conservation constraint:

$$\frac{\partial V^{el}}{\partial q_i} = 0 \quad (5)$$

$$\nabla_{\mu_i} V^{el} = 0 \quad (6)$$

In the models presented in this report the charges are constrained to give charge neutral molecules

$$\sum_{i=1}^{N_\beta} q_{i\beta} = 0 \quad (7)$$

where  $N_\beta$  is the number of charges in molecule  $\beta$ . One can equivalently express the charge constraint implicitly by transforming to a set of generalized charge coordinates.<sup>21</sup>

The polarizable model used for the polypeptide (PFF) comes from the work of Kaminski et al.<sup>1</sup> The model places fixed partial charges on all atomic positions and on massless virtual sites representing the lone pairs of the oxygen and sulfur atoms. The electrostatic parameters are fit from gas-phase electronic structure calculations<sup>25</sup> using density-functional theory (DFT) with the B3LYP method<sup>38,39</sup> and the cc-pVTZ(-f) basis set. The choice of basis set, which does not include diffuse functions, is based on evidence that including contributions from such functions in gas-phase DFT calculations results in an overpolarization of the parametrized model in the condensed phase.<sup>26,40,41</sup> The effect is likely a result of an energetic cost, hindering polarization, that results from Pauli repulsion between neighboring molecules in the condensed phase.<sup>42</sup> The polarizabilities of the atomic sites are parametrized by a series of electrostatic perturbations, using dipolar probes applied to the target molecule. The resulting change in the electrostatic potential is measured at a set of grid points outside the van der Waals surface of the molecule. Polarizabilities ( $\alpha_i$ ) are chosen to minimize deviations from the DFT calculation. The fixed charges and the parameters  $\gamma_i$  are chosen to best approximate the electrostatic potential from the unperturbed DFT calculation.

Stretching and bending energies for PFF are retained from the OPLS-AA force field,<sup>3</sup> while the torsional energy is reparametrized.<sup>1</sup> Further details can be found in the respective references. The electrostatic energy consists of a system of fixed point charges and point polarizable dipoles described by eq 3. The (1,2) and (1,3) interactions are omitted owing to the breakdown of the bare Coulomb potential at such short intersite distances. No intermolecular screening of the Coulomb potential is included in the original model formulation. Short-range repulsion and dispersion is represented by a Lennard-Jones function

$$U_{ij} = \sum_{i \neq j} 4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] f_{ij} \quad (8)$$

where we apply the geometric sum rule ( $\sigma_{ij} = (\sigma_i \sigma_j)^{1/2}$  and  $\epsilon_{ij} = (\epsilon_i \epsilon_j)^{1/2}$ ) for the interaction between particle  $i$  and  $j$ . The function  $f_{ij}$  is a scaling factor equal to zero for particles connected by a valence bond or angle, set to 0.5 for intramolecular 1,4 interactions and is 1.0 for all other pairs. The Lennard-Jones parameters are derived from ab initio dimer energies of organic compound analogues of the residues and from the OPLS-AA force field.

We employ three commonly used water models to solvate the polypeptide. A fixed charge TIP4P<sup>43</sup> model (for the fixed charge OPLS-AA protein<sup>3,44</sup>) and two polarizable water models, TIP4P-FQ<sup>19</sup> fluctuating charge model and an inducible point dipole model, RPOL.<sup>8</sup> All three models employ an interacting Lennard-Jones site placed on the oxygen atom.

**Table 1:** Screening Radius Applied to Selected Dipole Sites on the Polarizable Peptide Molecule in Solvated TIP4P-FQ Simulations (Å)

residue	(*) site of screened dipole	screening radius
glutamic acid	-C*O—O	2.5
aspartic acid	-C*O—O	2.5
aspartic acid	-CO*—O*	1.8
methionine	-S—C*H <sub>3</sub>	2.0
tyrosine	-C*OH—	2.5

Intermolecular interactions between electrostatic sites is described by the bare Coulomb potential (see eq 4). The TIP4P-FQ model includes an intramolecular interaction between the charges within the molecule that is parametrized along with the other electrostatic parameters empirically. The RPOL model places point polarizable dipoles on the oxygen and hydrogen atoms and omits intramolecular electrostatic coupling.

In practice using this energy function for the simulation of the TIP4P-FQ model with the polarizable protein results in a polarization catastrophe where the electrostatic variables between an interacting molecular pair mutually enhance to infinite polarization. The polarization catastrophe arose from interactions between TIP4P-FQ water and specific residues on the protein (see Table 1). This problem is a direct consequence of the point charge/dipole approximation to the electron charge distribution. For point polarizable models the pair interaction energy diverges at intersite distances proportional to the molecules' polarizability<sup>1/3</sup>. A similar phenomena is found in fluctuating charge and combined fluctuating charge/polarizable dipole models. A simple illustration is a pair of isotropic interacting point inducible dipoles where the singularity occurs at<sup>45</sup>

$$r_{critical} = (4\alpha_i\alpha_j)^{1/6} \quad (9)$$

Models that incorporate a large molecular polarizability by variable point charges/dipoles may have values for  $r_{critical}$  that approach physically relevant interaction distances. When this is the case, it is necessary to replace the Coulomb potential with a more accurate representation of the true potential at small interaction distances. A scaling factor may be applied to the Coulomb function, or more rigorously a screening function<sup>46,26</sup> may be used to effectively smear the point multipole and more accurately represent the potential of an electron charge distribution.

One should note the difficulty in accurately modeling polarization with a simple potential function may lead to an exaggerated polarizability and therefore a large  $r_{critical}$ .<sup>42,40</sup> It is interesting to point out that the polarizable RPOL water model did not result in a polarization catastrophe when used to solvate the polarizable protein. A more detailed study is necessary to definitively resolve whether the catastrophe in the TIP4P-FQ solvation model is a product of an unrealistically large polarizability of the water and protein or the use of the Coulombic coupling between the electrostatic sites. Providing evidence for the former explanation, a recent study<sup>47</sup> has shown a significant overpolarization response for hydrogen bonding configurations in the neat fluid using the

TIP4P-FQ model, which may result from exaggerated components of the molecular polarizability along the molecular plane. Settling potential problems with the TIP4P-FQ water model is a prerequisite to dealing with the source of the polarization catastrophe in our solvated protein simulation. This work is ongoing in our lab. In lieu of a satisfactory resolution on the TIP4P-FQ water model, we adopt measures to dampen the polarization response between specific residues and TIP4P-FQ water by applying a screening function for close range intermolecular interactions. The cubic spline  $f(x)$  is chosen such that  $f(0) = 1$ ,  $f'(0) = 0$ ,  $f(1) = 1$  and  $f'(1) = -1, -2, -3$ , correspond to the value of the functions  $1/x$ ,  $1/x^2$ ,  $1/x^3$ , respectively, at  $x = 1$ .<sup>26</sup> A “screening radius” is applied to specific dipole sites on the protein which affects the charge-dipole interactions. A summary of the sites and the respective screening radii is given in Table 1. The Coulomb potential  $u(r) = 1/r^2$  is replaced with

$$u(r) = \frac{1}{s^2} f(r/s) \quad (10)$$

when  $r < s$  and  $s$  is the sum of the screening radii on the pair of interacting sites.

**B. Polarization and MD. 1. P3M Ewald with Dipoles.** The models studied in this report include point dipoles in addition to point charges to describe the polarizable system; therefore, in addition to interactions between charges the electrostatic potential needs to describe the interactions between charge–dipole and dipole–dipole sites.

The Ewald sum for evaluating the Coulombic energy for a system of point charges has been extended to a system of multipoles by Smith.<sup>7</sup> The electrostatic energy for a periodic system of point charges and point dipoles is

$$E_{elec} = \frac{1}{2} \sum_{\mathbf{n}} \sum_i \sum_j' \frac{(q_i + \mu_i \cdot \nabla_i)(q_j + \mu_j \cdot \nabla_j)}{|\mathbf{r}_{ij} + \mathbf{n} \cdot \mathbf{L}|} \quad (11)$$

The Ewald sum with the metallic boundary condition follows from ref 7

$$E^r = \frac{1}{2} \sum_i \sum_{j \neq i} (q_i + \mu_i \cdot \nabla_i)(q_j + \mu_j \cdot \nabla_j) \frac{\text{erfc}(\eta r_{ij})}{r_{ij}} \quad (12)$$

$$E^k = \frac{1}{2} \sum_i \sum_j \sum_{\mathbf{k} \neq 0} \frac{1}{V} \frac{4\pi}{k^2} (q_i + \mu_i \cdot \nabla_i)(q_j + \mu_j \cdot \nabla_j) e^{-k^2/4\eta^2} e^{-i\mathbf{k} \cdot \mathbf{r}_{ij}} \quad (13)$$

$$E^s = -\frac{\eta}{\sqrt{\pi}} \sum_i \left( q_i^2 + \frac{2\eta^2}{3} |\mu_i|^2 \right) \quad (14)$$

where  $\mathbf{L}$  is the unit cell dimensions and the Ewald splitting parameter  $\eta$  modulates the relative weight of  $E^r$  and  $E^k$  to the total potential energy. For excluded  $i,j$  pairs we subtract  $1/r_{ij}$  which is equivalent to replacing  $\text{erfc}(\eta r_{ij})/r_{ij}$  in  $E^r$  with  $-\text{erf}(\eta r_{ij})/r_{ij}$ . Mesh based approximations to the regular Ewald sum discretize space on a regular grid reducing the Fourier series transforms in  $E^k$  to finite Fourier transforms which can be evaluated by fast Fourier transform (FFT)



algorithms.<sup>48,49</sup> The Fourier space portion of the electrostatic energy is evaluated using a suitable extension of the P3M Ewald (P3ME) method,<sup>33,32</sup> as discussed below. As is the case for a system of point charges the real space part and self-energy remain unchanged when using this generalized P3ME method.

Recently Toukmaji et al.<sup>36</sup> have extended the SPME method to include dipole–charge and dipole–dipole interactions. We adopt a similar extension to the P3ME method for the applications presented in this study. Following a similar formalism to that used in ref 36 we define a charge dipole array and follow the same four step procedure for generating the forces outlined in refs 32 and 33 for a fixed charge system:

$$Q_i = q_i + \mu_i \cdot \nabla_i \quad (15)$$

We assign this array to a grid using the same spline function ( $P$  is the order of the spline) applied to the fixed charge system

$$\begin{aligned} \rho_M(\mathbf{r}_p) &= \frac{1}{V} \sum_{h=1}^N \sum_{p=0}^P Q_i W^p(\mathbf{r}_p - \mathbf{r}_i) \\ &= \frac{1}{V} \sum_{h=1}^N \sum_{p=0}^P (q_i W^p(\mathbf{r}_p - \mathbf{r}_i) + \mu_i \cdot \nabla_i W^p(\mathbf{r}_p - \mathbf{r}_i)) \end{aligned} \quad (16)$$

where  $\mathbf{r}_p$  are the positions of the grid sites and  $M$  denotes the spatial grid. In a similar spirit to the “analytic” scheme for evaluating the spatial gradients in the field calculation discussed in ref 32 the dipole gradients are evaluated by analytic differentiation of the spline function,  $W^p$ , which can be factorized into a product of its Cartesian components.<sup>31</sup> The spline functions up to order  $P = 7$  are tabulated in ref 32. After assignment we apply a forward FFT to get the Fourier space charge/dipole density

$$\hat{\rho}_M(\mathbf{k}) = V_c \sum_{\mathbf{r}_p \in M} \rho_M(\mathbf{r}_p) e^{-i\mathbf{k} \cdot \mathbf{r}_p} \quad (17)$$

and the reciprocal space potential is

$$\hat{\phi}_M^k(\mathbf{k}) = \hat{G}(\mathbf{k}) \hat{\rho}_M(\mathbf{k}) \quad (18)$$

where the wave vectors are periodic with values  $\mathbf{k} = 2\pi\mathbf{m}/L$  and  $\mathbf{m}$  is an integer vector with values between  $-N_p/2 \leq \mathbf{m} < N_p/2$  and  $N_p$  is the number of grid points along each Cartesian axis. We use the optimized function  $\hat{G}$  corresponding to that derived for a system of point charges.<sup>31,33</sup>

$$\hat{G}_{opt}(\mathbf{k}) = \frac{\sum_n \hat{\mathbf{D}}(\mathbf{k}_n) \cdot \hat{\mathbf{R}}(\mathbf{k}_n) [\hat{W}(\mathbf{k}_n)/V_c]^2}{\sum_n [\hat{W}(\mathbf{k}_n)/V_c]^2 \sum_n |\hat{\mathbf{D}}(\mathbf{k}_n)|^2 [\hat{W}(\mathbf{k}_n)/V_c]^2} \quad (19)$$

The function  $\mathbf{k}_n = \mathbf{k} + 2\pi\mathbf{n}/h$ . The parameter  $V_c$  is the volume of the grid cell and  $h$  is the cell width. The function  $\hat{\mathbf{D}}(\mathbf{k}_n)$  is the Fourier transform of the differential operator and is  $i\mathbf{k}$  in this study. The function  $\hat{\mathbf{R}}(\mathbf{k}_n)$  is the Fourier transform of the true reference force

$$\hat{\mathbf{R}}(\mathbf{k}) = -i\mathbf{k} \frac{4\pi}{k^2} e^{-k^2/4\eta^2} \quad (20)$$

In principle an optimized function,  $\hat{G}$ , can be found that corresponds specifically to charge–charge, charge–dipole, and dipole–dipole interactions. However this is not a practical solution for an efficient algorithm, and the possible gains in accuracy are negligible. This is discussed further in Appendix I.

Using an inverse FFT we get the potential on the real space grid.

$$\phi_M^k(\mathbf{r}_p) = \frac{1}{V_{\mathbf{m}=-N_p/2}^{N_p/2-1}} \sum \hat{\phi}_M^k(\mathbf{k}) e^{i\mathbf{k} \cdot \mathbf{r}_p} \quad (21)$$

The polarizable models studied require the potential, field, and force to propagate the fluctuating charges, polarizable dipoles, and nuclei, respectively (see eq 26). To evaluate the forces and the field it is necessary to interpolate the potential back to the particles. This is done in a similar way to the analytic differentiation method of ref 32. The resultant equations are

$$\phi^k(\mathbf{r}_i) = \sum_{p=0}^P \phi_M^k(\mathbf{r}_p) W^p(\mathbf{r}_i - \mathbf{r}_p) \quad (22)$$

$$\begin{aligned} \mathbf{E}(\mathbf{r}_i) &= -\nabla_i \phi^k(\mathbf{r}_i) \\ &= -\sum_{p=0}^P \phi_M^k(\mathbf{r}_p) \nabla_i W^p(\mathbf{r}_i - \mathbf{r}_p) \end{aligned} \quad (23)$$

$$\begin{aligned} \mathbf{F}(\mathbf{r}_i) &= (q_i + \mu_i \cdot \nabla_i) \mathbf{E}(\mathbf{r}_i) \\ &= -\sum_{p=0}^P \{q_i \phi_M^k(\mathbf{r}_p) \nabla_i W^p(\mathbf{r}_i - \mathbf{r}_p) + \phi_M^k(\mathbf{r}_p) \mu_i \cdot \nabla_i (\nabla_i W^p(\mathbf{r}_i - \mathbf{r}_p))\} \end{aligned} \quad (24)$$

This method requires only 2 FFT’s in order to evaluate the forces.

**2. Extended Lagrangian Formalism for Polarization.** Eqs 5 and 6 lead to a set of coupled linear equations in the total electric field and the total potential at site  $i$  which can be solved iteratively until self-consistency is achieved. Upward of six iterative calculations of the electric field and potential may be necessary at each step in the molecular dynamics simulation in order to conserve energy.<sup>36</sup> Considering that the field calculation is the most expensive portion of a molecular dynamics simulation, the calculation of polarization in this manner leads to at least a 6-fold increase in the computational complexity. For polarizable force fields an alternative to iteratively solving for the electric degrees of freedom at each time step in the simulation is an approximate method similar in spirit to the Car-Parrinello ab initio MD method.<sup>10,19,37,50</sup> This method treats the electric degrees of freedom as dynamical variables by defining an extended Lagrangian for the equations of motion

$$L = \frac{1}{2} \sum_i^N [m_i \dot{r}_i^2 + m_q \dot{q}_i^2 + m_\mu \dot{\mu}_i^2] - V - \sum_\beta \lambda_\beta \sum_{i=1}^{N_\beta} q_{i\beta} \quad (25)$$

where  $V$  is the total potential energy,  $\lambda_\beta$  is a Lagrange multiplier necessary to satisfy charge neutrality on each

molecule, and  $N_\beta$  is the number of charge sites on molecule  $\beta$ . The charge/dipole dynamics are fictitious and serve solely to keep these variables near the minimum energy, in a computationally efficient manner. The corresponding Lagrangian equations of motion are<sup>19,10</sup>

$$m_q \ddot{q}_j = -\frac{1}{N_\beta} \sum_{i=1}^{N_\beta} (\phi_j - \phi_i)$$

$$m_\mu \ddot{\mu}_j = -\alpha_j^{-1} \mu_j + \mathbf{E}_j \quad (26)$$

where  $\phi_j$  and  $\mathbf{E}_j$  are the electrostatic potential and total field on site  $j$  respectively and  $\sum_i$  is over all fluctuating charge sites on the molecule containing site  $j$ .

The extended Lagrangian approximation requires that the fictitious dynamics run approximately adiabatically for the duration of the simulation. This is satisfied with an appropriately small choice of  $m_{q/\mu}$ . If the frequency  $\omega_q$  and  $\omega_\mu$  of the electric variables is sufficiently larger than the fastest nuclear frequency, the degree of thermal coupling between the fictitious and real dynamics will be small. The weaker this coupling, the longer the simulation will progress with the variable charges and dipoles remaining near the minimum energy surface. However, values for the fictitious frequency that are very large require a time step in the simulation that is proportionally small resulting in a computationally costly simulation. It is necessary to strike a balance in the selection of the fictitious frequency parameters between the degree of thermal coupling to the nuclear dynamic bath and the computational cost of the simulation. A typical time step for integrating a biomolecular system is on the order of a femtosecond. Simulations that require a time step significantly smaller than this will be prohibitively slow. The frequency for the electrostatic variables is approximately  $\omega_q = (J^0/m_q)^{1/2}$  for the fluctuating charges and  $\omega_\mu = 1/(\alpha m_\mu)^{1/2}$  for the polarizable dipoles with an isotropic polarizability. We choose a charge and dipole mass ( $m_q, m_\mu$ ) such that  $\omega_q$  and  $\omega_\mu = 53, 333 \text{ cm}^{-1}$ . By applying a stability analysis to the leapfrog algorithm, Hockney and Eastwood<sup>31</sup> have derived that the relationship between the time step used in the finite difference integration algorithm and the largest frequency in the system is  $\omega_{\max} \Delta t_{\max} = 2$ . The charge/dipole frequency used in this study ( $\omega_{\max} = 53, 333 \text{ cm}^{-1}$ ) gives a  $\Delta t_{\max} = 1.25 \text{ fs}$ . However, the motion of the charge/dipole variables are coupled by the electric potential/field leading to frequency modes in the system larger than  $53, 333 \text{ cm}^{-1}$ . In practice we found a system dependent time step ranging between 0.75 fs and 1 fs was necessary for stable integration of the equations of motion. The extended Lagrangian temperature remained below 0.5 K for all the polarizable model simulations indicating these degrees of freedom remain near the minimum energy surface. Thus on the time scale of the extended Lagrangian simulations (approximately 300 ps between iterative minimization of the electrostatic configuration) there is no appreciable transfer of energy to the fictitious degrees of freedom. For very long simulations (much greater than 300 ps), not tested in this study, it may be necessary to systematically minimize the charge/dipole configuration in order to ensure a more accurate representa-

tion of the model system. Considering the infrequency that such a minimization would be needed, the added computational cost is essentially zero. Previous studies have found it necessary to apply thermostats and in some cases a restraining potential, to keep the charge/dipole variables near the minimum energy surface.<sup>20,24,40</sup> It is not clear how such an approach affects the dynamics of the system. A faithful representation of the nuclear dynamics requires that the energy flow to the fictitious degrees of freedom is negligible. This condition can be transparently satisfied when using a constant NVE simulation with the extended Lagrangian protocol applied in this report, by monitoring the extended Lagrangian temperature as a function of simulation time.

### III. Application: Water Solvated BPTI

Bovine pancreatic trypsin inhibitor (BPTI) has been used as a benchmark for force field simulations for some time,<sup>51,52</sup> and for this reason is our choice of protein for this study. We compare equilibrium simulations of BPTI using a newly developed polarizable force field for polypeptides<sup>1</sup> (PFF), solvated in TIP4P-FQ and RPOL water, to a fixed charge representation of BPTI using the OPLS-AA force field with fixed charge TIP4P water. Regarding the accuracy of their model, Kaminski et al.<sup>1</sup> make note that their model is a first generation attempt at developing a quantitatively accurate force field for biomolecular simulation which includes polarization. Further development and testing in the condensed phase are prerequisites to refining the model. Incorporating an efficient methodology for large-scale polarizable condensed phase simulation, as is applied in this study, is a step toward that goal. As such we are interested in a stable efficient simulation for long time scales and a reasonable representation of native state stability.

**A. Simulation Procedure.** The simulation procedure is as follows. The starting structure, obtained from the 4PTI structure in the Brookhaven Protein Data Bank,<sup>53</sup> included a protein monomer and 60 water molecules. Hydrogen atoms were added using the MAESTRO software package.<sup>54</sup> Six counter chloride ions were added using GENION,<sup>55</sup> to neutralize the system. BPTI was then solvated in a 60 Å cubic unit cell of water generated from a preequilibrated box of neat TIP4P. Removing water molecules that overlap with the protein left 6377 water molecules in the system. The equilibration and production simulations were performed using the program SIM developed in our lab.<sup>56</sup> The initial equilibration procedure used nonpolarizable force fields and proceeded as follows. Using the OPLS-AA force field and keeping the protein structure fixed, the water solvent was equilibrated at constant temperature (298 K) and pressure (1 atm) for 20 ps using Nose-Hoover chain thermostats<sup>57</sup> and Andersen-Hoover type barostats,<sup>58</sup> giving a cubic unit cell of length  $L = 58.8 \text{ Å}$ . To generate the starting configurations for the polarizable model production simulations, an additional 10 ps of simulation at constant volume and constant temperature (NVT) with a fixed protein preceded an additional 50 ps of simulation in NVT allowing the protein and water to relax. The production simulations were run in the microcanonical ensemble using the velocity Verlet algorithm for 2 ns with a 1 fs time step (0.75 fs for

PFF/TIP4P-FQ). A further 500 ps of equilibration was found to be necessary leaving the final 1.5 ns for analysis. The RATTLE<sup>59</sup> constraint algorithm was used to keep the water molecular geometry rigid, and the bonds between the protein heavy atoms and hydrogens were held fixed. The Ewald parameters for the simulations were  $\eta = 0.37 \text{ \AA}^{-1}$ , a spherical truncation of the real space potential at  $R_{\text{cut}} = 10 \text{ \AA}$  applied to the electrostatic site positions, a grid spacing of  $0.75 \text{ \AA}$ , and an assignment order  $P = 6$ . The minimum energy electrostatic configuration is solved iteratively at the outset of the simulation. The extended Lagrangian method was used to propagate these variables during the simulation. The availability of computing time dictated the duration of each simulation segment (65–600 ps). To generate the 2 ns trajectories the simulation segments were run in sequential order using the final nuclear configuration of the previous segment as the initial configuration for the following segment. For computational convenience the variable charges/dipoles are iteratively minimized at the outset of each segment. It should be noted that the minimizations are not motivated by a drift of these variables from the minimum energy surface. For example the increase in the extended Lagrangian temperature is only  $\approx 0.01 \text{ K}$  for the 315 ps segment using PFF/TIP4P-FQ.

**B. Results and Discussion.** *1. Efficiency.* To analyze the speedup from adopting the extended Lagrangian formalism and the P3ME approximation we compare to a regular Ewald calculation of the electrostatic energy with an iterative solution to the electrostatic variables. For a consistent level of accuracy with the P3ME simulation (relative rms Force  $\approx 10^{-5}$ ) an efficient parameter set<sup>60</sup> for the regular Ewald sum corresponds to  $\eta = 0.25 \text{ \AA}^{-1}$ , a spherical truncation of the real space potential at  $R_{\text{cut}} = 15 \text{ \AA}$  and of the wave vectors  $k_{\text{max}} = 14 \text{ \AA}^{-1}$ . The P3ME method is approximately 8 times faster than Ewald in evaluating the forces. An iterative simulation requires greater than six iterations of the field calculation per time step.<sup>36</sup> In contrast, the overhead in using the extended Lagrangian formalism is less than a factor of 1.1. A conservative estimate of the computational gain is therefore on the order of 40 for an extended Lagrangian/P3ME polarizable simulation compared to an iterative/Ewald scheme even for this relatively small protein. For a similarly sized system of neat RPOL water, Toukmaji et al.<sup>36</sup> have reported a speedup of 100 using an extended Lagrangian/SPME simulation. The difference lies not in the relative efficiency of SPME vs P3M Ewald, which is very similar, but in the choice of suboptimal parameters for the regular Ewald calculation used in the comparison between SPME and Ewald. To compare the computational cost of a polarizable simulation to fixed charge models we present timing data for the execution of one molecular dynamics time step. Remarkably the fully polarizable model is only a factor of 1.23–1.45 more expensive than the fully fixed charge model (the ratios are relative to the fixed charge simulation). The bulk of the computational effort comes from the evaluation of the electrostatic interactions. The TIP4P-FQ model requires no new interactions and thus requires the smaller computational effort (1.23). Using truncation methods and the regular Ewald sum, models that incorporate inducible

**Table 2:** Time Averaged RMSD between the Average NMR Structure and the Average Simulation Structure<sup>a</sup> (Å)

model	C <sub>α</sub>	heavy atoms
OPLS-AA/TIP4P	0.7	1.17
PFF/TIP4P-FQ	0.85	1.47
PFF/RPOL	0.91	1.39

<sup>a</sup> Terminal residues are excluded from the comparison.

dipoles have been shown to be a factor of 2 more expensive than analogous fixed charge models.<sup>10,26</sup> However in the P3ME formalism there is no additional expense in the evaluation of the FFT's when using a model with shared charge/dipole sites compared to charges alone. For this reason the RPOL polarizable dipole solvation model gives the improved scaling (1.45).

The molecular dynamics program used in this study is benchmarked against a popular MD program in order to put the preceding timing experiments into a familiar context. SIM shows comparable computational speed relative to AMBER 7.<sup>4</sup> Our SIM molecular dynamics program is approximately 1.3 times slower than AMBER 7 on a 2.4 Ghz Pentium IV processor for a 23558 atom sized system of fixed charges.

*2. Simulation Accuracy.* Energy conservation in the NVE ensemble is one measure of how faithfully our simulation represents the model Hamiltonian given the P3ME and extended Lagrangian approximations employed. The total energy fluctuations,  $\Delta V$ , provide a measure of the energy conservation and the total energy drift

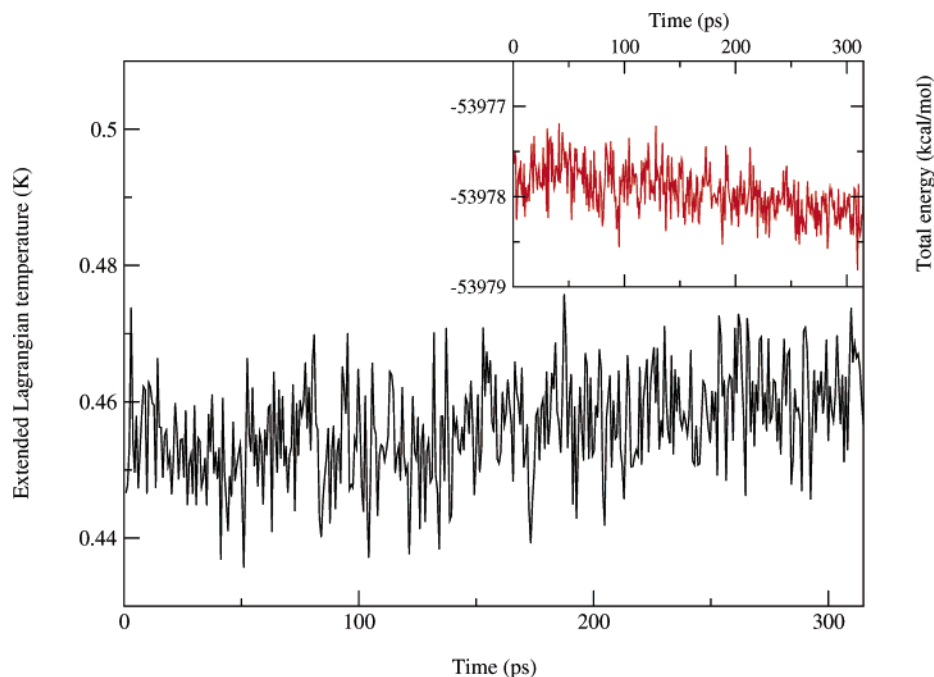
$$\Delta V = \frac{1}{N_T} \sum_{i=1}^{N_T} \left| \frac{V_0 - V_i}{V_0} \right| \quad (27)$$

where  $V_i$  is the total energy at step  $i$ ,  $V_0$  is the initial energy under the extended Lagrangian dynamics, and  $N_T$  is the total number of time-steps. This quantity has been interpreted as a reasonable measure of accuracy,<sup>61–63</sup> and a value of  $\Delta V \leq 0.003$ , i.e.,  $\log(\Delta V) < -2.5$ , gives an acceptable numerical accuracy. Another parameter that measures the simulation accuracy is the ratio of root-mean-square fluctuations between the total energy ( $\Delta V_{\text{rms}}$ ) and the kinetic energy ( $\Delta KE_{\text{rms}}$ )

$$R = \frac{\Delta V_{\text{rms}}}{\Delta KE_{\text{rms}}} \quad (28)$$

A value of  $R < 0.05$  has been correlated with good energy conservation.<sup>64</sup> In Table 3 we present these parameters for the various simulations. The largest values for  $\log(\Delta V)$  and  $R$  are  $-4.66$  and  $0.004$ , respectively, far less than the acceptable minimum level of accuracy, indicating very good energy conservation. One should note the excellent performance of our simulations indicates a conservative array of P3ME and integration parameters. A less strict tolerance may be sufficient and result in a faster simulation.

The accuracy of the configurational trajectory for the fluctuating charge and polarizable dipoles is related to the extended Lagrangian temperature. If the electrostatic variables begin to drift from the potential energy minimum, the corresponding extended Lagrangian temperature will begin to increase. The extended Lagrangian temperature remains



**Figure 1.** Trajectory of the total energy and extended Lagrangian temperature for PFF/TIP4P-FQ simulation. The 315 ps duration is the longest simulation segment used between minimizations for this model combination. The temperature stays approximately constant for the duration indicating little thermal coupling with the nuclei on this time scale. The total energy for the same trajectory is also plotted showing fluctuations about a consistent value for the simulation length.

**Table 3:** Summary of the Simulation Accuracy

model	$\log(\Delta V)^a$	$R^b$	extended Lagrangian T (K) <sup>c</sup>	$q_{\text{rmsd}}$ (e)	$\mu_{\text{rmsd}}$ (Debye)
OPLS-AA/TIP4P	-5.21	0.004	N/A	N/A	N/A
PFF/TIP4P-FQ	-4.96	0.003	0.5	0.002	0.012
PFF/RPOL	-5.30	0.004	0.1	N/A	0.003

<sup>a</sup>  $\log(\Delta V)$  measures fluctuations about a reference energy near the beginning of the simulation. <sup>b</sup>  $R$  measures the ratio of root-mean-square fluctuations between the total and kinetic energy. Accurate fixed charge simulations have been correlated with a value of  $\log(\Delta V) < -2.5$  and a value of  $R < 0.05$ . <sup>c</sup> The extended Lagrangian temperature measures the degree of thermal coupling between the electrostatic variables and the nuclear variables. Typically temperatures less than 6 K have been correlated with a good representation of the potential ground state.

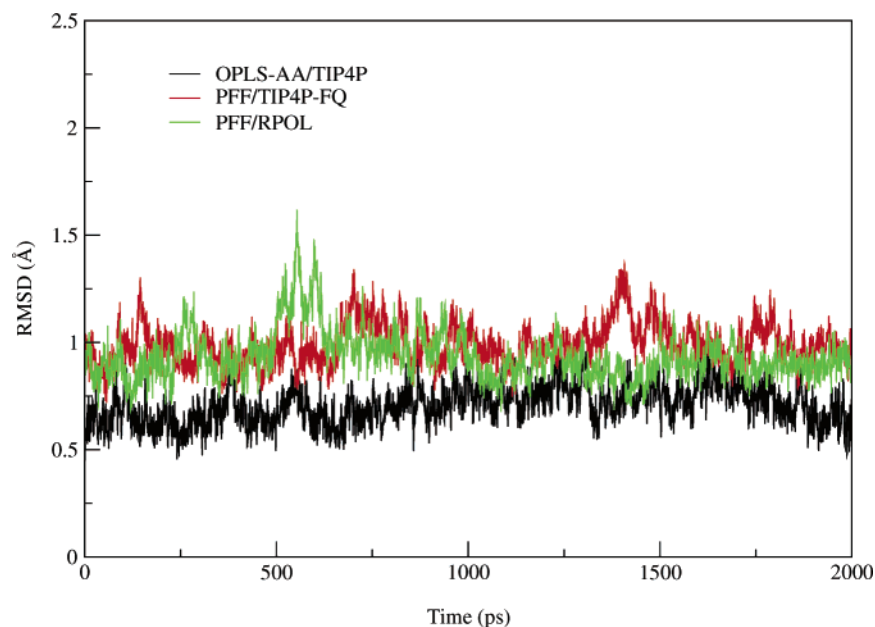
near the value at the beginning of the polarizable simulations never rising above 0.5 K (see Table 3), which indicates these variables remain near the potential energy minimum for the duration of the simulation (for the TIP4P-FQ model a temperature  $< 6$  K has been correlated with a good representation of the minimum energy electrostatic configuration<sup>19</sup>). Figure 1 shows the extended Lagrangian temperature for the longest simulation segment between iterative minimizations (315 ps) using the TIP4P-FQ model. Table 3 also shows the root-mean-square deviations of the electrostatic variables from the minimum energy configuration for the last nuclear configuration corresponding to the largest simulation segment between minimizations (400 ps for RPOL). The deviations are small providing further evidence the generated trajectories are representative of the minimum energy surface.

**3. Force Field Accuracy.** How well simulations represent the native structure of the protein provides a coarse measure-

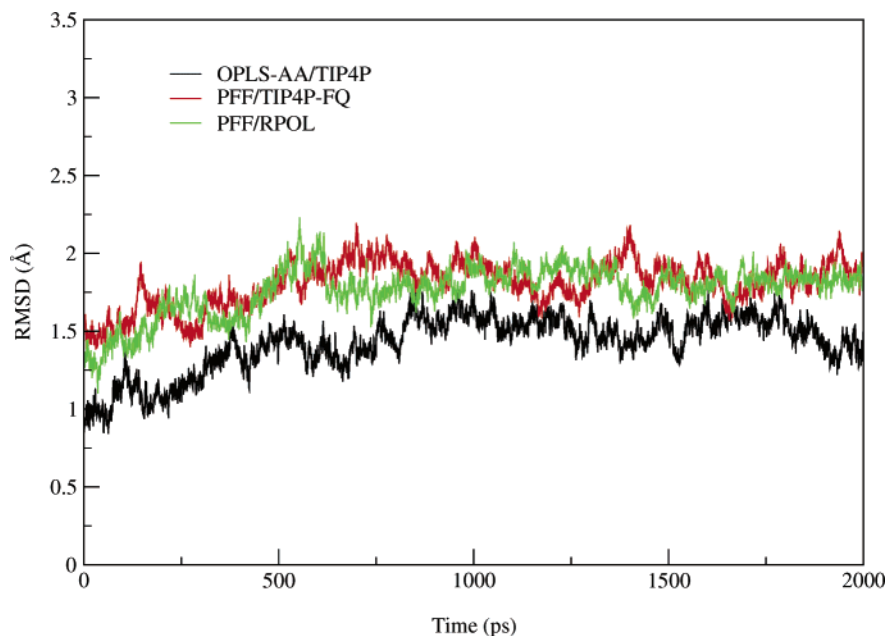
ment of the quality of the polarizable force field models. Assuming the model simulation begins in the real native state, a poor representation will lead to large deviations as it relaxes to the erroneous native state of the model. Measuring small deviations from experiment over long simulation periods (nanoseconds) is a positive indication for the model. In this study the experimental crystal structure is used to represent the native state. NMR experiments of BPTI in liquid water<sup>65</sup> have shown small root-mean-square deviations (RMSD) between the average NMR structure and the crystal structure (RMSD=0.85 Å) indicating the native protein structure does not change significantly between the liquid and crystal. We present the RMS deviations of the simulated peptide backbones relative to the experimental crystal structure for 2 ns simulations in Figure 2. The terminal residues show large fluctuations in the liquid water simulations and are not included in the RMSD analysis in line with the results from the NMR experiments.<sup>65</sup> The polarizable force field simulations are similar to the fixed charge model with the average RMS deviation being  $\approx 0.8$  Å for all the models. For a direct comparison to an experimental solvated structure, time averaged RMSD values between simulation and the average NMR structure are evaluated and summarized in Table 2. The differences between the polarizable and nonpolarizable models remain small and within the experimental error.

We also investigate the RMS deviations of the heavy atoms of the protein (see Figure 3). The results show close similarity between the polarizable and nonpolarizable model simulations. It is interesting to note that equilibrium of the protein including the side chains requires an additional 500 ps of simulation as evidenced by the slope at the start of the





**Figure 2.** Root-mean-square deviation of C $\alpha$  atoms for the simulation structures from the experimental crystal structure as a function of simulation time. Terminal residues, which show large fluctuations from NMR experiments as well as simulation, are not included in this analysis. All models do a reasonable job representing the protein native state for the 2 ns duration of the simulation.



**Figure 3.** Root-mean-square deviation, including all heavy atoms, for the simulation structures from the experimental crystal structure as a function of simulation time.

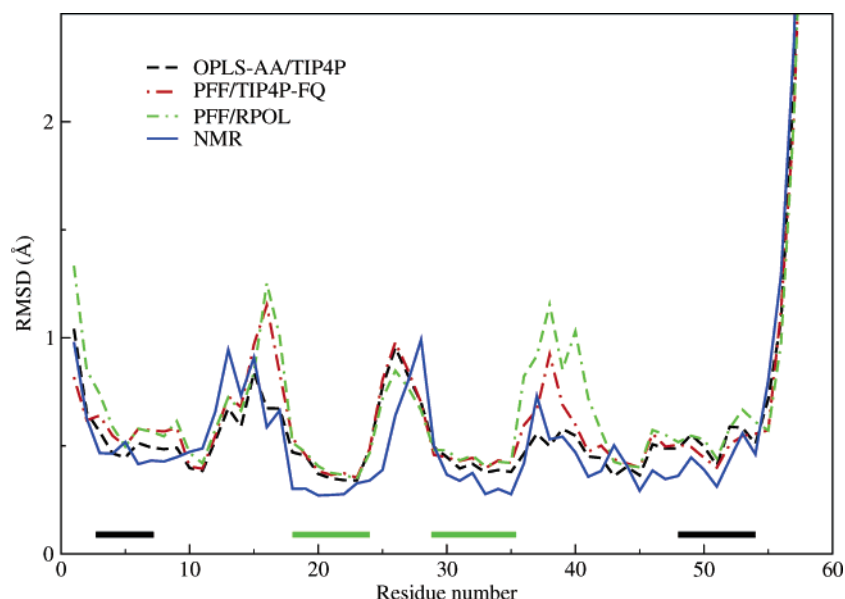
simulations in Figure 3. Figure 4 shows the time averaged RMS deviation corresponding to each residue in the protein from simulation and NMR experiment relative to their average structures. The results are again similar for all three simulations.

Considering the small structural deviations between simulation and the protein crystal structure are similar to that from the nonpolarizable force field one can only conclude that like the fixed charge models the polarizable simulations are a reasonable representation of this water/BPTI system. The results are similar for averaged RMSD values between simulation and the average NMR structure (see Table 2). A

more sensitive experimental probe is necessary to resolve the relative accuracy of the nonpolarizable OPLS-AA simulation and the proposed polarizable solvation models studied in this report.

#### IV. Conclusions

We have presented a computationally efficient and accurate methodology for the simulation of large polarizable systems using a combination of fluctuating charges and polarizable dipoles. The method requires only a modest overhead relative to nonpolarizable force fields and the simulations run stably for 2 ns. The method does not need thermostats or the



**Figure 4.** Time averaged RMSD over main chain atoms of individual residues from the respective average simulation and NMR structure. The  $\beta$ -strands are marked in green, and the  $\alpha$ -helices are marked in dark gray. The residue dependence of the protein fluctuations in the liquid show correspondence between the simulation models and the NMR experiment.

imposition of restraint potentials, which may obscure the dynamics of the system, to keep the electrostatic degrees of freedom near the minimum energy surface. Thus, the method allows a reliable representation of the system dynamics at constant energy and constant volume. Our tests of a newly developed polarizable protein force field combined with the TIP4P-FQ and RPOL models for water gave promising levels of accuracy compared to the experimental structures. However, a more detailed study is necessary to resolve the relative accuracy of this protein polarizable model to that of fixed charge polypeptide models. Experiments that probe the hydrogen bonding environment of these solvated peptides, such as time-resolved infrared spectroscopy, are a promising tool for evaluating the quality of biomolecular force fields.<sup>47</sup> Although the structural quantities investigated in this report show little to differentiate the polarizable and nonpolarizable simulations, one should not conclude that the simulations are similar. Significant differences between the nonpolarizable and polarizable simulations are found in the hydrogen bonding patterns of the protein structure and in the structural and dynamic properties of the solvent surrounding the protein. This will be presented in a following publication.<sup>66</sup> One should also keep in mind that the polarizable peptide model is in the first stage of development. Further refinement of the model, by including condensed phase data into the parametrization, is now possible and will certainly lead to more accurate polarizable models.

**Acknowledgment.** The authors thank Dr. Tom Young and Dr. Shenglong Wang for helpful discussions. This work was supported by a grant to B.J. Berne from NSF (CHE-03-16896), to Richard A. Friesner from NIH (GM52018) and by a grant of computer time from the EMSL at Pacific Northwest National Laboratories.

## V. Appendix: Optimized Coulomb Propagator for a System of Point Charges and Point Dipoles

Following the procedure of Hockney and Eastwood<sup>31</sup> one can derive optimized Coulomb propagators on a discretized space for charge–dipole and dipole–dipole interactions. Defining the mean square difference between the force calculated on the grid,  $\mathbf{F}$ , and the true reference force for the continuous space problem,  $\mathbf{R}$ , to be

$$Q = \frac{1}{V_h} \int_{V_h} d^3\mathbf{r}_1 \int_V d^3\mathbf{r} |\mathbf{F}(\mathbf{r}; \mathbf{r}_1) - \mathbf{R}(\mathbf{r})|^2 \quad (29)$$

The Fourier space representations of the reference forces for the interacting charge/dipole sites are

$$\begin{aligned} \hat{\mathbf{R}}_{qq} &= q_i q_j i\mathbf{k} \frac{4\pi}{k^2} e^{-k^2/4\eta^2} \\ \hat{\mathbf{R}}_{qp} &= q_i (\mathbf{k} \cdot \boldsymbol{\mu}_j) i\mathbf{k} \frac{4\pi}{k^2} e^{-k^2/4\eta^2} \\ \hat{\mathbf{R}}_{pq} &= -(\mathbf{k} \cdot \boldsymbol{\mu}_i) q_j i\mathbf{k} \frac{4\pi}{k^2} e^{-k^2/4\eta^2} \\ \hat{\mathbf{R}}_{pp} &= -(\mathbf{k} \cdot \boldsymbol{\mu}_i)(\mathbf{k} \cdot \boldsymbol{\mu}_j) i\mathbf{k} \frac{4\pi}{k^2} e^{-k^2/4\eta^2} \end{aligned} \quad (30)$$

where  $qq$  denotes the force between two charges,  $qp$  denotes the force on charge  $q$  from dipole  $p$ ,  $pq$  denotes the force on  $p$  from  $q$ , and  $pp$  denotes the force between two dipoles. The calculated force from the discretized space is

$$\begin{aligned} \hat{\mathbf{F}}_{qq} &= q_i q_j \hat{\mathbf{U}} \hat{\mathbf{D}} \hat{\mathbf{G}} \sum_n \hat{\mathbf{U}} e^{i(\mathbf{k} - \mathbf{k}_n) \cdot \mathbf{r}_1} \\ \hat{\mathbf{F}}_{qp} &= q_i \hat{\mathbf{U}} \hat{\mathbf{D}} \hat{\mathbf{G}} \sum_n \hat{\mathbf{U}} (\mathbf{k}_n \cdot \boldsymbol{\mu}_j) e^{i(\mathbf{k} - \mathbf{k}_n) \cdot \mathbf{r}_1} \end{aligned}$$

$$\hat{\mathbf{F}}_{pq} = -(\mathbf{i}\mathbf{k} \cdot \boldsymbol{\mu}_i) q_j \hat{\mathbf{U}} \hat{\mathbf{D}} \hat{\mathbf{G}} \sum_n \hat{U} e^{i(\mathbf{k} - \mathbf{k}_n) \cdot \mathbf{r}_1}$$

$$\hat{\mathbf{F}}_{pp} = (\mathbf{k} \cdot \boldsymbol{\mu}_i) \hat{\mathbf{U}} \hat{\mathbf{D}} \hat{\mathbf{G}} \sum_n \hat{U} (\mathbf{k}_n \cdot \boldsymbol{\mu}_j) e^{i(\mathbf{k} - \mathbf{k}_n) \cdot \mathbf{r}_1} \quad (31)$$

where  $\hat{U} = \hat{W}/V_h$  and  $\hat{\mathbf{D}}$  depends on the method of potential differentiation for the field gradient. The wave vector  $\mathbf{k} = 2\pi\mathbf{m}/\mathbf{L}$  and  $\mathbf{k}_n = \mathbf{k} + 2\pi\mathbf{n}/\mathbf{h}$ . Minimizing the functional derivative of  $\hat{Q}$  with respect to  $\hat{G}$  we can get the optimized functional parameter  $\hat{G}$  corresponding to this set of interactions. However since  $i\mathbf{k}$  is not periodic in the alias sum over  $\mathbf{n}$ , the Fourier transform of the dipole gradient remains within the alias sum (see page 274 of ref 31 for details of the derivation), and the parameters  $\hat{G}_{pq}$ ,  $\hat{G}_{qp}$ , and  $\hat{G}_{pp}$  remain explicit functions of the particular dipole site,  $\mu$ . Obviously this is not a tenable solution. A finite difference approximation for differentiating the dipole gradients is periodic, and the result reduces to eq 19 for all interacting pairs. Of course as was shown in ref 33 the difference in  $\hat{G}$  between keeping  $i\mathbf{k}$  within the alias sum and factoring it out all but disappears for the assignment orders and grid densities used in this study, which means to an excellent approximation,  $\hat{G}_{pq} = \hat{G}_{qp} = \hat{G}_{pp}$ .

## References

- (1) Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A.; Cao, Y. X.; Murphy, R. B.; Zhou, R.; Halgren, T. A. *J. Comput. Chem.* **2002**, *23*, 1515.
- (2) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fisher, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E., III; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, D.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586.
- (3) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225.
- (4) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.
- (5) Stone, A. *Chem. Phys. Lett.* **1981**, *83*, 233.
- (6) Sagui, C.; Pedersen, L. G.; Darden, T. A. *J. Chem. Phys.* **2004**, *120*, 73.
- (7) Smith, W. *CCP5 Newsletter* **1998**, *46*, 18.
- (8) Smith, D. E.; Dang, L. X. *J. Chem. Phys.* **1994**, *100*, 3757.
- (9) Dang, L. X.; Chang, T. *J. Chem. Phys.* **1997**, *106*, 8149.
- (10) van Belle, D.; Froeyen, M.; Lippens, G.; Wodak, S. J. *Mol. Phys.* **1992**, *77*, 239.
- (11) Mountain, R. D. *J. Chem. Phys.* **1995**, *103*, 3084.
- (12) Wallqvist, A.; Berne, B. J. *J. Phys. Chem.* **1993**, *97*, 13841.
- (13) Bernardo, D. N.; Ding, Y.; Krogh-Jespersen, K.; Levy, R. M. *J. Phys. Chem.* **1994**, *98*, 4180.
- (14) Ding, Y.; Bernardo, D. N.; Krogh-Jespersen, K.; Levy, R. M. *J. Phys. Chem.* **1995**, *99*, 11575.
- (15) Meng, E. C.; Caldwell, J. W.; Kollman, P. A. *J. Phys. Chem.* **1996**, *100*, 2367.
- (16) Gao, J.; Habibollahzadeh, D.; Shao, L. *J. Phys. Chem.* **1995**, *99*, 16460.
- (17) Cieplak, P.; Caldwell, J.; Kollman, P. *J. Comput. Chem.* **2001**, *22*, 1048.
- (18) Kaminski, G. A.; Jorgensen, W. L. *J. Chem. Soc., Perkin. Trans. 2* **1999**, *11*, 2365.
- (19) Rick, S. W.; Stuart, S. J.; Berne, B. J. *J. Chem. Phys.* **1994**, *101*, 6141.
- (20) Rick, S. W.; Berne, B. J. *J. Am. Chem. Soc.* **1996**, *118*, 672.
- (21) Liu, Y.-P.; Kim, K.; Berne, B. J.; Friesner, R. A.; Rick, S. W. *J. Chem. Phys.* **1998**, *108*, 4739.
- (22) Stuart, S. J.; Berne, B. J. *J. Phys. Chem.* **1996**, *100*, 11934.
- (23) Banks, J. L.; Kaminski, G. A.; Zhou, R.; Mainz, D. T.; Berne, B. J.; Friesner, R. A. *J. Chem. Phys.* **1999**, *110*, 741.
- (24) Patel, S.; Brooks, C. *J. Comput. Chem.* **2004**, *25*, 1.
- (25) Stern, H. A.; Kaminski, G. A.; Banks, J. L.; Zhou, R.; Berne, B. J.; Friesner, R. A. *J. Phys. Chem. B* **1999**, *103*, 4730.
- (26) Stern, H. A.; Berne, B. J.; Friesner, R. A. *J. Chem. Phys.* **2001**, *115*, 2237.
- (27) York, D.; Darden, T.; Pedersen, L. *J. Chem. Phys.* **1993**, *99*, 8345.
- (28) Belhadj, M.; Alper, H. E.; Levy, R. M. *Chem. Phys. Lett.* **1991**, *179*, 13.
- (29) Ewald, P. *Ann. Phys.* **1921**, *64*, 253.
- (30) Essmann, U.; Perera, L.; Berkowitz, M.; Darden, T.; Lee, H.; Pedersen, L. *J. Chem. Phys.* **1995**, *103*, 8577.
- (31) Hockney, R. W.; Eastwood, J. W. *Computer Simulations Using Particles*; IOP: Bristol, 1988.
- (32) Deserno, M.; Holm, C. *J. Chem. Phys.* **1998**, *109*, 7678.
- (33) Zhou, R.; Harder, E.; Xu, H.; Berne, B. J. *J. Chem. Phys.* **2001**, *115*, 2348.
- (34) Sagui, C.; Darden, T. *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 155.
- (35) Procacci, P.; Darden, T.; Marchi, M. *J. Chem. Phys.* **1996**, *100*, 10464.
- (36) Toukmaji, A.; Sagui, C.; Board, J.; Darden, T. *J. Chem. Phys.* **2000**, *113* (24), 10913.
- (37) Sprik, M.; Klein, M. L. *J. Chem. Phys.* **1988**, *89*, 7556.
- (38) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098.
- (39) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.
- (40) Lamoureux, G.; MacKerell, A. D.; Roux, B. *J. Chem. Phys.* **2003**, *119*, 5185.
- (41) Patel, S.; MacKerell, A. D., Jr.; Brooks, C. L. *J. Comput. Chem.* **2004**, *25*, 1504.
- (42) Morita, A.; Kato, S. *J. Chem. Phys.* **1999**, *110*, 11987.
- (43) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. *J. Chem. Phys.* **1983**, *79*, 926.
- (44) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **2001**, *105*, 6474.
- (45) Bottcher, C. F. J. *Theory of Electric Polarization, Vol I*, 2nd ed.; Elsevier: 1973.
- (46) Thole, B. T. *Chem. Phys.* **1981**, *59*, 341.
- (47) Harder, E.; Eaves, J.; Tokmakoff, A.; Berne, B. J. manuscript in preparation.

- (48) Brigham, E. O.; Morrow, R. E. *IEEE Spectrum* **1967**, 4, 63.
- (49) Frigo, M.; Johnson, S. G. *ICASS conference proceedings* **1998**, 3, 1381.
- (50) Car, R.; Parrinello, M. *Phys. Rev. Lett.* **1985**, 55, 2471.
- (51) York, D.; Wlodawer, A.; Pedersen, L. G.; Darden, T. *Proc. Natl. Acad. Sci.* **1994**, 91, 8715.
- (52) Daggett, V.; Levitt, M. *Annu. Rev. Biophys. Biomol. Struct.* **1993**, 22, 353.
- (53) Deisenhofer, J.; Steigemann, W. *Acta Crystallogr. B* **1975**, 31, 238.
- (54) v. 6.0, M. Schrödinger, Inc.; Portland OR, 2003.
- (55) van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hünenberger, P. H.; Krueger, P.; Mark, A.; Scott, W.; Tironi, I. *Biomolecular Simulation: The GROMOS96 Manual and User Guide*; Hochschulverlag AG/ETH Zurich, 1996.
- (56) Stern, H.; Rittner, F.; Pavese, M.; Harder, E.; Xu, H.; Kim, B. *SIM: Molecular Dynamics Simulation Program*.
- (57) Tuckerman, M. E.; Martyna, G. J.; Klein, M. L. *J. Chem. Phys.* **1992**, 97, 2635.
- (58) Martyna, G. J.; Tobias, D. J.; Klein, M. L. *J. Chem. Phys.* **1994**, 101, 4177.
- (59) Andersen, H. C. *J. Comput. Phys.* **1983**, 52, 24.
- (60) Petersen, H. G. *J. Chem. Phys.* **1995**, 103.
- (61) Watanabe, M.; Karplus, M. *J. Chem. Phys.* **1993**, 99, 8063.
- (62) Zhou, R.; Berne, B. J. *J. Chem. Phys.* **1995**, 103.
- (63) Figueirido, F.; Zhou, R.; Levy, R.; Berne, B. J. *J. Chem. Phys.* **1997**, 106, 9835.
- (64) van Gunsteren, W. F.; Berendsen, H. J. C. *Mol. Phys.* **1977**, 34 (5), 1311.
- (65) Berndt, K. D.; Guntert, P.; Orbons, L. P. M.; Wuthrich, K. *J. Mol. Biol.* **1992**, 227, 757.
- (66) Kim, B.; Young, T.; Harder, E.; Berne, B. J.; Friesner, R. A. manuscript in preparation.

CT049914S