



Published in final edited form as:

*J Chem Theory Comput.* 2013 June 11; 9(6): 2826–2837. doi:10.1021/ct400116g.

## Matching of additive and polarizable force fields for multiscale condensed phase simulations

Christopher M. Baker<sup>1</sup> and Robert B. Best<sup>\*,1,2</sup>

<sup>1</sup>University of Cambridge, Department of Chemistry, Lensfield Road, Cambridge, CB2 1EW, UK

<sup>2</sup>Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892-0520, U.S.A

### Abstract

Inclusion of electronic polarization effects is one of the key aspects in which the accuracy of current biomolecular force fields may be improved. The principal drawback of such approaches is the computational cost, which typically ranges from 3 – 10 times that of the equivalent additive model, and may be greater for more sophisticated treatments of polarization or other many-body effects. Here, we present a multiscale approach which may be used to enhance the sampling in simulations with polarizable models, by using the additive model as a tool to explore configuration space. We use a method based on information theory to determine the charges for an additive model that has optimal overlap with the polarizable one, and we demonstrate the feasibility of enhancing sampling via a hybrid replica exchange scheme for several model systems. An additional advantage is that, in the process, we obtain a systematic method for deriving charges for an additive model that will be the natural complement to its polarizable parent. The additive charges are found by an effective coarse-graining of the polarizable force field, rather than by *ad hoc* procedures.

### Keywords

Force Field; Polarizable; Relative Entropy; Coarse-graining; Multiscale Simulation

### Introduction

Additive force fields have been very successful in reproducing condensed phase properties of macromolecules.<sup>1</sup> Nonetheless, it is generally recognized that the form of the energy function used in such simulations is inadequate to reproduce the molecular potential energy surface to within  $k_B T$  accuracy under standard conditions.<sup>2,3</sup> A major deficiency is the use of fixed point charges to represent intermolecular electrostatic interactions, neglecting atomic polarizability. The approximation of using fixed charges also leads to ambiguities in which charges should be chosen. For example, if charges are obtained by fitting to a molecular electrostatic potential,<sup>4</sup> the results will invariably depend on the internal degrees of freedom of the molecule.<sup>5–7</sup> This is generally solved in an *ad hoc* way by fitting the charges simultaneously to a number of configurations,<sup>8,9</sup> but it is not clear what relative weights should be used in the fit. A second problem is that a condensed phase aqueous

\*Corresponding author: robertbe@helix.nih.gov.

Supporting Information. Five supporting figures showing convergence of  $S_{\text{rel}}$  with number of configurations used, the effect of different charge restraint forces, radial distribution functions for cyclohexane, potential energy surface for cyclohexane, energy difference distributions comparing the SWM4-NDP water model to several additive water models; and seven tables giving force field parameters, charge variations and results of cyclohexane simulations.

environment will tend to increase the overall polarization of a molecule.<sup>2</sup> This effect has been approximately included in various ways, including the use of an implicit solvent model in the calculation of the electrostatic potential,<sup>10</sup> explicit consideration of interactions with water molecules,<sup>11–13</sup> or by using Hartee-Fock theory to determine a gas phase electrostatic potential (RESP fitting),<sup>8,14</sup> on the grounds that this tends to result in overpolarization due to the neglect of electron correlation. In summary, charges for additive force fields are usually determined by reasonable, but rather *ad hoc*, procedures. Choosing the best possible charges for additive models is very important, however, as such models are the only computationally practical approach to many problems. These issues are driving a renewed interest in methods for determining charges in additive force fields. For example, a recent study has proposed an iterative scheme using two alternating components: (i) simulations with the current additive charges in the condensed phase and (ii) derivation of new additive charges using a RESP scheme, but with selected solvent molecules from the simulation explicitly included.<sup>15</sup> Our work presents an alternative approach to this problem.

The shortcomings of additive force fields have motivated the development of more sophisticated functional forms including many-body interactions,<sup>16,17</sup> in particular polarizable force fields.<sup>18,19</sup> The most sophisticated polarizable models aim to reproduce the energy surface to within 0.5 kcal/mol ( $\sim k_B T$  at 300 K), but come at a significant computational cost. Simulations with polarizable force fields typically run from  $\sim 2$ – $3$  (CHARMM Drude model in NAMD,<sup>20</sup> Amber ff02<sup>21</sup>) to  $\sim 10$  (AMOEBA<sup>3,22</sup>) times slower than simulations with additive force fields (and can be much slower for still more sophisticated models). For simulations of biomolecular dynamics, which are already challenging with additive force fields due to the rugged energy landscape,<sup>23,24</sup> this additional cost can be prohibitive. It would therefore be desirable to combine the advantages of polarizable and additive models, using additive models to improve sampling, and polarizable models for greater accuracy.

Luchi et al. have previously shown that, for water, coarse-graining based on force-matching<sup>25</sup> can be used to produce a non-polarizable model that accurately reproduces bulk properties obtained from a fully polarizable water model.<sup>26</sup> In this paper, we propose to adopt a similar approach, viewing the derivation of charges for an additive force field as a coarse-graining problem. We start from the assumption that we have a polarizable force field of the desired accuracy available, noting that, in principle at least, determining electrostatic parameters for polarizable models from gas phase quantum chemical calculations should be less ambiguous than for additive models.<sup>3</sup> We then choose the set of fixed charges which gives maximum overlap between the configurational distributions obtained with additive and polarizable models, for simulations in an aqueous environment. This approach has two important outcomes: firstly, it provides a well-posed definition of the optimal set of fixed charges. Secondly, it should also be the best choice for running “resolution exchange” simulations<sup>27</sup> in which polarizable and additive force fields are employed in a replica exchange scheme, allowing the goal of enhanced sampling with the additive force field to be realized. We present the application of the method to pure water, solutions of acetamide in water, and cyclohexane, and demonstrate its application in resolution exchange simulations of water and cyclohexane. We also discuss the issue of degeneracy of the solutions, how it can be overcome in practice and its relation to the variation in charge parameters within and across existing force fields.

## Methods

### Theory

Our starting point is the relative entropy formalism for comparing distributions, which has been proposed as the basis for a coarse-graining procedure by Shell.<sup>28–32</sup> In this approach, a

“relative entropy”  $S_{\text{rel}}$  (or Kullback-Leibler divergence<sup>33</sup>), measures the similarity between a “target” simulation  $T$  and a “model” simulation  $M$  with configuration distribution functions  $p_T(x)$  and  $p_M(x)$  respectively, where  $x$  represents a point in configuration space.

$$S_{\text{rel}} = \int p_T(x) \ln \frac{p_T(x)}{p_M(x)} dx \quad (1)$$

Shell has proposed a gradient-based approach for variationally optimizing the parameters  $\lambda = \lambda_1, \lambda_2, \dots$  of the model energy function by means of a gradient minimization of the relative entropy<sup>30,31</sup> Using a Newton-Raphson approach for the minimization, the parameter update step is given by:

$$\lambda^{k+1} = \lambda^k - \chi \mathbf{H}^{-1} \nabla_{\lambda} S_{\text{rel}} \quad (2)$$

where  $\chi$  is a parameter chosen to ensure stability of the algorithm, and the gradient  $\nabla_{\lambda} S_{\text{rel}}$  and the Hessian  $\mathbf{H}$  of the relative entropy with respect of the parameters can be evaluated by averages over the target and model ensembles as follows:

$$\nabla_{\lambda} S_{\text{rel}} = \beta \left\langle \frac{\partial U_M}{\partial \lambda} \right\rangle_T - \beta \left\langle \frac{\partial U_M}{\partial \lambda} \right\rangle_M \quad (3)$$

$$\mathbf{H}_{ij} = \beta \left\langle \frac{\partial^2 U_M}{\partial \lambda_i \partial \lambda_j} \right\rangle_T - \beta \left\langle \frac{\partial^2 U_M}{\partial \lambda_i \partial \lambda_j} \right\rangle_M + \beta^2 \left\langle \frac{\partial U_M}{\partial \lambda_i} \frac{\partial U_M}{\partial \lambda_j} \right\rangle_M - \beta^2 \left\langle \frac{\partial U_M}{\partial \lambda_i} \right\rangle_M \left\langle \frac{\partial U_M}{\partial \lambda_j} \right\rangle_M \quad (4)$$

The drawback of using equations 3 and 4 is that they require the calculation of properties averaged over both the target and model ensembles. Since the model is a function of the parameters  $\lambda$ , this means that a simulation using the latest version of the parameters must be performed at every step in the minimization. Chaimovich and Shell<sup>31</sup> have, however, noted that  $S_{\text{rel}}$  can be calculated via Equation 5:

$$S_{\text{rel}} = \ln \langle e^{\Delta - \langle \Delta \rangle} \rangle_T \quad (5)$$

Where  $\Delta = \beta(U_T - U_M)$  and  $U_T$  and  $U_M$  are the energies of configurations from the target ensemble evaluated using the target and model potentials, respectively. This expression allows us to calculate  $S_{\text{rel}}$  exactly without the averaging over the model ensemble, but the average over the exponential makes it prone to statistical error. If, however, we make the assumption that the distribution of  $\Delta$  is Gaussian, as would be expected from a linear response approximation, we can calculate  $S_{\text{rel}}$  via Equation 6,<sup>31</sup> reducing the statistical error.

$$S_{\text{rel}} \approx \frac{1}{2} \text{Var}_T[\beta(U_T - U_M)] \quad (6)$$

Using Equation 6 to calculate  $S_{\text{rel}}$ , we then evaluate  $\nabla_{\lambda} S_{\text{rel}}$  and  $\mathbf{H}$  directly using finite differences, and use these quantities to iterate Equation 2 until  $\lambda$  converges to the value at which the relative entropy is a minimum. As an alternative, we have also used a Monte-Carlo-based method to optimize  $S_{\text{rel}}$  calculated via Equation 6: this is discussed further below.

While this is the method that we used in practice, we also note that the derivatives in Equations 3 and 4 can be obtained using a combination of energy evaluations in which the atomic partial charges in the atom-centered model are set to either unity or zero. This approach is outlined in detail in the Supporting Information. We note that in this case the problem of having to resample the model distribution can be potentially alleviated by reweighting for a certain number of steps until the overlap between model and target distribution becomes low enough that resampling is needed.<sup>34</sup> In principle, a weighted histogram method could even be used to combine information from multiple parameter sets, to suppress oscillations in parameter values.<sup>35</sup> However, as we found the finite difference method to be adequate, we did not pursue these methods here.

## Simulation Methods

Sets of “target” configurations to be used in the relative entropy optimization were generated by running molecular dynamics (MD) simulations using the CHARMM Drude polarizable force field<sup>2</sup> in the program CHARMM.<sup>36</sup> Three such simulations were performed: one of liquid water (a cubic box of 250 molecules); one of a single ACEM molecule in a box of 250 water molecules, and one of a single CHEX molecule in a box of 250 water molecules. All simulations used the SWM4-NDP water model,<sup>37</sup> with ACEM and CHEX parameters from refs<sup>38</sup> and <sup>39</sup>, respectively. The simulations were performed using periodic boundary conditions, with the equations of motion integrated using a velocity Verlet integrator<sup>40</sup> and Drude particles treated via an extended Lagrangian double thermostat formalism.<sup>41</sup> A mass of 0.4 amu was transferred from each heavy atom to its Drude particle and the second thermostat was applied to maintain the temperature of the Drude oscillators at 1K. A Nosé-Hoover thermostat<sup>42,43</sup> (relaxation time 1 ps) was applied to all atoms and a modified Anderson-Hoover barostat<sup>44</sup> (relaxation time 0.1 ps) was used to maintain a constant pressure of 1 atm. Covalent bonds to hydrogen were constrained using the SHAKE algorithm<sup>45</sup> and Lennard-Jones interactions were treated explicitly to 12 Å, with atom-based switch-smoothing applied over the range of 10–12 Å; a long-range correction<sup>46</sup> to the Lennard-Jones interactions was also included. Electrostatic interactions were treated using particle mesh Ewald<sup>47</sup> (PME) summation with a 32 point grid in each dimension, a coupling parameter of 0.34 and a sixth-order spline for mesh interpolation. Each simulation was run for 1.5 ns, starting from a pre-equilibrated box, with a 1 fs timestep and coordinates saved every 100 steps. For water, 1,500 of the saved configurations were used in the optimization of the relative entropy; for ACEM and CHEX, 10,000 and 5,000 of the saved configurations were used in the relative entropy optimization, respectively. These values were chosen to ensure the accurate computation of the relative entropy in all cases (Figure S1).

Radial distribution functions for each of the molecules, with the polarizable force field, were calculated from the simulations described above. Molecular volumes and enthalpies of vaporization for water were calculated using a simulation protocol identical to that described above, but with ten simulations, each of 150 ps duration, each starting from the same configuration but with velocities initiated using different random seeds. Final properties were then averaged over all ten simulations, with uncertainties calculated as the standard error of the mean. Diffusion coefficients,  $D$ , and dielectric constants,  $\epsilon$ , were obtained from four independent simulations, each of 5 ns with the final 4.5 ns used for analysis. Diffusion coefficients were calculated using the Einstein relation,<sup>48</sup> while dielectric constants were calculated as described in Ref. <sup>49</sup>, with final values averaged over all four simulations and uncertainties estimated as the standard error of the mean. High-frequency optical dielectric constants,  $\epsilon_{\text{inf}}$ , were estimated using the Clausius-Mossotti equation. Hydration free energies for all molecules were calculated using free energy perturbation<sup>50</sup> (FEP) via the staged protocol of Deng and Roux,<sup>51</sup> in a matter identical to that described previously.<sup>52</sup> Final

values are obtained as averages over five independent calculations, with uncertainties calculated as the standard error of the mean.

Following optimization of charge parameters using the relative entropy formalism, additional simulations were performed to calculate physical properties with the new, non-polarizable models. The simulation protocol used was the same as that described above, but without the inclusion of any Drude particles, which in turn permits the use of a 2 fs timestep. For the sake of comparison, analogous simulations were also performed using well-established, non-polarizable models: the TIP3P<sup>53</sup> water model; the TIP4P-Ew<sup>54</sup> water model; ACEM from CGenFF<sup>13</sup> (the CHARMM General Force Field), and CHEX from the CHARMM sugar force field.<sup>39</sup>

As a further test of the optimized parameters, a number of homo- and hetero-dimeric interactions with water were considered. In these calculations, the water geometry was fixed to the SWM4-NDP/TIP3P geometry while the solute was optimized within the force field of interest. With the geometries of both molecules fixed, the water molecule was then placed in a specific orientation relative to the solute molecule, and the separation of the molecules increased in increments of 0.1 Å with the total energy of the complex evaluated at every increment. The interaction energy was then defined as the difference between the energy of the complex and the sum of the energies of the two individual molecules.

Modifications to the CHARMM source code allowed us to run replica exchange molecular dynamics (REMD) simulations<sup>55</sup> involving the Drude polarizable force field, and to exchange configurations between polarizable and non-polarizable force fields. The principal challenge in implementing such a scheme is that the polarizable and non-polarizable force fields each contain different numbers of particles: there are no Drude particles present in the non-polarizable simulations. In practice, we dealt with this by including additional zero-mass, non-interacting dummy atoms in the non-polarizable simulations. We added these dummy particles to the additive CHARMM topology so that their positions in the final PSF were equivalent to the positions of the Drude particles in the polarizable PSF, ensuring that replica exchange would swap additive dummy atoms with polarizable Drude particles. Then, when an exchange was attempted, the following procedure was adopted: (i) before the attempted swap, the Drude particles were optimized in the polarizable model, and the energy re-evaluated; (ii) atomic coordinates were swapped between the systems, (iii) the positions of the Drude particles were again optimized via a full self-consistent field calculation in the polarizable force field, before the energies were re-evaluated for both polarizable and additive models, (iv) the move was decided according to the exchange criterion below. Thus, the energies for both sets of coordinates were for adiabatic Drude particles. The exchange criteria are given in Equation 7, where  $E_a(\mathbf{X}_a)$  represents the energy of the coordinates from replica b evaluated using the potential function from replica a;  $\Delta_i = E_i(\mathbf{X}_i) - E_j(\mathbf{X}_i)$ ;  $\Delta_j = E_i(\mathbf{X}_j) - E_j(\mathbf{X}_j)$ , and  $\beta_i = 1/k_B T_i$ .

$$P_{i \leftrightarrow j} = \begin{cases} 1 & (\beta_i \Delta_i - \beta_j \Delta_j) \leq 0 \\ e^{(\beta_i \Delta_i - \beta_j \Delta_j)} & (\beta_i \Delta_i - \beta_j \Delta_j) > 0 \end{cases} \quad (7)$$

The modified CHARMM code was used to perform replica exchange simulations that coupled polarizable and non-polarizable simulations for liquid water and liquid cyclohexane. The simulation protocols were the same as those described above, with replica exchange typically attempted every 100 steps. Velocities were exchanged together with coordinates and scaled by the factor  $(T_{\text{new}}/T_{\text{old}})^{1/2}$ , where  $T_{\text{new}}$  and  $T_{\text{old}}$  are respectively the temperatures after and before exchange.<sup>55</sup>

## Results and Discussion

### 1. Additive Force Field Optimization

As our reference polarizable potential, we have used the CHARMM Drude force field.<sup>37–39,52,56–68</sup> The development of this force field ultimately targets its application to biomolecular simulation,<sup>2</sup> and it has been implemented within the programs CHARMM<sup>36</sup> and NAMD.<sup>20</sup> Our work is implemented in CHARMM, where benchmark simulations show that the polarizable force field is ~5 times more costly than a comparable additive model. During development of this force field, a number of small molecules have been explicitly parameterized, and we have used the CHARMM Drude models of water;<sup>37</sup> acetamide<sup>61</sup> (ACEM), and cyclohexane<sup>39</sup> (CHEX) (Figure 1).

**1.1 Water**—Water represents a natural starting point in any force field development project due to its ubiquity in nature, role in influencing biomolecular conformation and interactions<sup>69–75</sup> and the large body of existing work on the development/evaluation of water models.<sup>76–80</sup> Furthermore, in this work, optimization of a water model must be the first step because calculations employing the optimized water model are required to optimize parameters for ACEM and CHEX. Our reference model, the SWM4-NDP water model,<sup>37</sup> is a four-site water model with three sites located at the nuclear positions and the fourth located on the line bisecting the H-O-H angle (Figure 1). The geometry of the three atomic sites (i.e. hydrogen and oxygen nuclei) is the same as that of the TIP3P water model.<sup>53</sup> In addition to these sites, there is also a “Drude” particle. In the CHARMM Drude polarizable force field, these additional particles are added to each heavy atom, with their charge determined by a polarizability parameter assigned to that atom. For water, this means that a single Drude particle is included, attached to the O atom. Our objective was to identify, using the relative entropy framework described above, an additive water model giving the best possible reproduction of the configuration distribution obtained using the SWM4-NDP model. We have therefore adopted a 4-site model in which charges are located on both the O atom and the virtual site. To generate the initial non-polarizable water model, all parameters from the SWM4-NDP model were transferred directly to a non-polarizable analogue, termed SWM4A-INI (initial 4-site additive water model) in which the polarizability was removed from the O atom, with all charges maintained. In producing our final additive model, only the point charges were optimized: all internal and van der Waals parameters were maintained at the values obtained from the SWM4-NDP model (shown in Tables S1–S5). In principle, variation of any force field parameter could play a role in improving the fit of the additive model, but the results are expected to be weakly dependent on parameters other than charges. Optimizing additional parameters from noisy data is also likely to be challenging. Due to symmetry and overall charge neutrality, the task of fitting the 4-site water is reduced to optimization of only two parameters, thus lowering potential degeneracy in the fit. The final additive model obtained from relative entropy minimization was termed SWM4A-OPT (optimized 4-site additive water model). The final charges obtained from the relative entropy optimization are shown in Table 1.

In Figure 2, we show the progress of the charge optimization. Using the gradient-based optimization discussed in Methods, the charges converge within five steps, accompanied by a significant decrease in the relative entropy. The relative entropies shown, along with the gradients used in the optimization, were calculated using Equation 6. As noted, this expression gives only an approximation of the true  $S_{\text{rel}}$  values obtained from Equation 5. To test whether the approximation was valid in this case, we also performed a 2-dimensional scan in  $q_{\text{O}}$  and  $q_{\text{LP}}$ , evaluating the relative entropy via both Equations 5 and 6 at every point. The numerical values of  $S_{\text{rel}}$  obtained with the two methods differed, but followed the same trends: both predicted the same minimum, and the overall correlation coefficient across 100

data points was 0.992. As a further test of the minimization approach based on Equation 6, we used Equation 3 to calculate the gradients at the final point of our optimization, confirming that it does represent a minimum. At this point, we also tested how much improvement might be expected from an optimization of the Lennard-Jones parameters on oxygen. However, while the initial charge optimization results in the relative entropy dropping from ~6000 to ~80, further optimization of the Lennard-Jones parameters results in a decrease to only ~66. Therefore, the improvement accessible by optimizing parameters other than charges was relatively small, and optimization of these parameters was subsequently neglected.

The final model, SWM4A-OPT, produced from this charge optimization yields physical properties in good agreement with those calculated using the reference polarizable force field. Figure 3 shows radial distribution functions calculated with various force field models. The SWM4A-INI model clearly provides a poor approximation to the SWM4-NDP model, while the SWM4A-OPT model is in very good agreement with its polarizable counterpart. It is also worth noting that the SWM4A-OPT model yields results closer to the polarizable model than does the widely-used TIP3P model, showing that optimizing charges via the relative entropy methods produces a model giving better agreement with the polarizable model than can be expected from an arbitrarily-chosen optimized water model. As a water model, SWM4A-OPT represents an improvement over TIP3P, which is known to be too weakly structured; the structural properties of the polarizable SWM4-NDP model are much closer to those inferred from experimental neutron scattering data.<sup>81</sup> Comparing to a well-optimized 4-site additive model, TIP4P-Ew<sup>54</sup>, also shows the SWM4A-OPT model to give results closer to those obtained with the polarizable force field. TIP4P-Ew is, in fact, more structured than the SWM4A-OPT model, and the pair distribution agrees better with the experimental data. The weaker structure of SWM4A-OPT is simply a consequence of the original polarizable SWM4-NDP model being too weakly structured.

Additional physical properties calculated with the various force fields are shown in Table 2. The SWM4A-OPT model performs poorly in reproducing gas-phase properties, but this is not unexpected. The gas phase is a very different dielectric environment from the aqueous liquid phase, and non-polarizable models simply cannot be expected to reproduce properties in both environments; this, of course, is the motivation for the development of explicitly polarizable models. Similarly the properties of the water dimers, which can be considered to be closer to the gas phase than the liquid phase, are not particularly well reproduced. In the liquid phase, however, the SWM4A-OPT model performs very well. Molecular volume and enthalpy of vaporization are explicitly considered as targets in the optimization of the CHARMM Drude polarizable force field, with the target being to reproduce the experimental values to within 2%,<sup>2</sup> and an error of magnitude up to 5% considered acceptable. Before comparing the values calculated for the heat of vaporization, however, we need to apply a correction due to the fact that the additive model of water, optimized for the condensed phase, effectively remains polarized even in the gas phase. That is, the energy associated with a condensed phase charge distribution is higher than the energy of the true gas phase charge distribution (an alternative viewpoint is that in the reverse process of condensation, the additive model incurs no energetic penalty for adopting the condensed phase charge distribution). Therefore, the heat of vaporization for an additive model is in general overestimated. An approximate correction for this effect, applicable to polarizable dipolar molecules, was derived by Berendsen *et al.*,<sup>82</sup> in which the polarization energy  $E_{\text{pol}}$  is given by:

$$E_{\text{pol}} = \frac{1}{8\pi\epsilon_0\alpha} N_A (\mu_l - \mu_g)^2 \quad (8)$$

In this expression,  $N_A$  is the Avogadro constant,  $\alpha$  is the polarizability,  $\epsilon_0$  the permittivity of free space and  $\mu_l$  and  $\mu_g$  the molecular dipole moments in the liquid and gas respectively. The same correction with opposite sign is applied to the solvation free energies,  $\Delta G_{\text{hyd}}$  (there is no difference in the gas phase entropy of the polarized and non-polarized molecules). To compute  $E_{\text{pol}}$  for the SWM4A-OPT model, we used  $\mu_g = 2.46$  D, the average dipole moment of SWM4A-OPT in the condensed phase,  $\mu_l = 1.85$  D, the experimental gas phase dipole moment, and the experimental gas phase polarizability  $\alpha = 1.44 \text{ \AA}^3$ . After including this correction, the SWM4A-OPT enthalpy of vaporization differs from the SWM4-NDP value by  $-2.1\%$ , with the molecular volume differing by  $2.8\%$ . Since neither of these properties were explicitly considered in the parameter optimization, this agreement is remarkable. Perhaps even more remarkable is the agreement in the dielectric constant and the diffusion coefficient. Both properties are well reproduced by the SWM4A-OPT model – at a level comparable to the TIP4P-Ew model, and much better than by the TIP3P model, whose large diffusion coefficient is a well-known shortcoming. After including the gas phase polarization correction, the hydration free energy obtained using the SWM4A-OPT model is also in reasonable agreement with that obtained from the SWM4-NDP model.

Note that while both the enthalpy of vaporization and the hydration free energy of SWM4A-OPT are in reasonable agreement with the original polarizable model *once polarization corrections have been applied*, the inability of the model to capture both gas phase and solution environments is an inherent problem of additive force fields; polarizable force fields should be the most generally applicable to capturing different environments, particularly where interfacial effects are relevant. In choosing to focus on the liquid phase, we are making the necessary compromise entailed by coarse-graining, but with the aim of doing so in an optimal way.

**1.2 Acetamide**—As an analogue of the protein backbone, the acetamide functionality is an important component of any biomolecular force field targeting the simulation of proteins. It also represents a more demanding test for any parameter optimization scheme. A 4-site model of water contains only two independent charge parameters. Acetamide, however, possesses seven distinct atom types, resulting in six charge parameters for optimization (the seventh is determined by the requirement that the total charge on the molecule must be zero). An additional complication arises because the reference polarizable force field includes charge sites not located at the nuclear positions. In several force fields, additional point charges are included to better represent the anisotropy in the electron density around certain atoms,<sup>83–86</sup> and the CHARMM Drude polarizable force field includes additional point charges on H bond accepting atoms to represent “lone pairs”. In acetamide, this results in two additional point charges attached to the amide O atom, located at the points where one would intuitively expect to find lone pairs. To generate the initial non-polarizable water model, all parameters from the Drude model were transferred directly to a non-polarizable analogue, termed ADD-INI (initial additive model); the polarizabilities were removed from the heavy atoms, and the lone-pairs were also removed, with their charges transferred to their parent O atom. All other charges were maintained at the values used in the polarizable force field. Charge parameters for the additive model were optimized to give a model termed ADD-OPT (optimized additive model); as in the case of water, all internal and van der Waals parameters were maintained at the values obtained from the Drude model<sup>61</sup> (Tables S1–S5). The reference simulations in this case were of a single acetamide molecule in a box of water; the acetamide parameters were optimized with the water molecules represented by the SWM4A-OPT model.

The charge optimization for acetamide proved to be less straightforward than that for water. Initial parameter optimization via the gradient-based method gave results in good agreement with the experimental properties, but we observed that initiating the optimization from



different starting points gave different sets of charges with different relative entropies (Figure 4A). These are classic symptoms of a multiple minimum optimization problem, which we sought to address by a Monte Carlo simulated annealing (MCSA)<sup>87,88</sup> approach in which the Gaussian approximation to the relative entropy was used as the energy. With this method we were indeed able to obtain solutions with similar relative entropies starting from different initial charges, but the final charges were still rather different, indicating that there are multiple degenerate solutions, or multiple sets of charges that all perform equally well when measured by relative entropy. From an algorithmic point of view this is unsatisfactory: we would like to obtain a well-defined set of charges in an unambiguous and reproducible fashion. We discuss the degeneracy further below, but in order to obtain a well-defined solution to this problem we have imposed restraints on the charges by optimizing the target function  $S'_{\text{rel}}$ , defined by:

$$S'_{\text{rel}} = S_{\text{rel}} + \frac{1}{2} k_q \sum_i (q_i - q_i^{\text{ref}})^2 \quad (9)$$

Here, we use a harmonic restraint with force constant  $k_q$  to keep each charge  $q_i$  close to some reference value  $q_i^{\text{ref}}$ . The natural choice for the reference charges is the set of initial charges derived from the polarizable model. With both the MCSA and charge restraints implemented, we were able to remove the dependency of the final charges on the initial starting configurations, as well as the degeneracy of the solution, without increasing the final, minimized value of the relative entropy (Figure 4B). Since the relative entropy is proportional to the negative log-likelihood of the target distribution given the model,<sup>28</sup> this is equivalent to a Gaussian prior for the charges. Clearly, one parameter which remains to be defined is the value of the restraint constant  $k_q$ ; we chose the value of  $k_q$  to be the largest that could be used in the Monte Carlo simulated annealing optimization without increasing the final value of the relative entropy. For acetamide, this value was found to be ~200 entropy units (Figure S2). The final charges obtained for ACEM using this approach are shown in Table 1, with properties calculated using the ADD-OPT model shown in Figure 5 and Table 3.

Acetamide-water RDFs are shown in Figure 5, and reveal good agreement between the polarizable and ADD-OPT models. Results from the ADD-INI model are noticeably worse than those with the ADD-OPT model, indicating that optimization of  $S_{\text{rel}}$  results in a significant improvement in the structural properties of the model, in spite of the relatively small changes in the magnitudes of the charges seen in Table 1. For reference, the RDFs obtained with the generic CGenFF force field are also shown in Figure 5, showing that the match with the polarizable force field is non-trivial. Other properties calculated with the ADD-OPT model (Table 3) also show acceptable agreement with those from the polarizable force field.

As alluded to above, in the absence of restraints on the charges, there is a certain amount of degeneracy in the optimal additive force fields for molecules with a large number of atom types such as acetamide. We characterize this degeneracy by noting that the negative log-likelihood of the target distribution  $T$  given the model charges  $\{q\}$  is<sup>28</sup>  $-\ln L(T|\{q\}) = nS_{\text{rel}}$  where  $n$  is the number of configurations used to compute  $S_{\text{rel}}$ . Adopting a Bayesian approach and assuming a uniform prior distribution of charges, we can determine the posterior distribution of charges given the target ensemble, since then  $L(\{q\}|T) \propto L(T|\{q\})$ . We sample the posterior distribution by running Metropolis Monte Carlo simulations in which  $nS_{\text{rel}}$  plays the role of the energy. We have characterized this distribution by computing the standard deviation of the charges from unrestrained Monte Carlo runs, obtaining values of 0.04–0.09 $e$ , depending on the atom type, although the variations will

clearly be correlated (Table S6). For comparison, we have also computed the standard deviation of the amide charges in non-prolyl amino acids in the Amber ff03 force field,<sup>10</sup> in which each residue is permitted to have a different set of charges on the backbone atoms, and also the standard deviation of the amide charges on the Alanine residue in a range of current force fields. Interestingly, we find that the standard deviation of charges across the different residues in the ff03 force field, as well as the standard deviation across force fields, is very similar to the variation in different solutions of our acetamide model. These results suggest that such variations could be at least partly explained by the degeneracy of solutions when fitting point charge force fields, which may support the use of a uniform charge scheme for all backbone atoms.<sup>8,89,90</sup>

In the development of the additive CHARMM force fields, charges are typically determined by considering heterodimeric interactions between the solute of interest and water molecules. The logic behind this is that it is a computationally cheap method that, to a first approximation, explicitly incorporates the effect of the environment into the parameterization,<sup>66</sup> ensuring that a correct balance between solute-water and water-water interactions is achieved.<sup>11</sup> This approach was originally adopted in the development of the CHARMM22 protein force field,<sup>11</sup> where the target energies and geometries were calculated *ab initio* at the HF/6–31G(d) level of theory, with interaction energies for neutral compounds multiplied by a factor of 1.16 and interaction distances offset by  $-0.2 \text{ \AA}$ . These latter corrections are introduced to account for deficiencies in the relatively low level of theory used (such as the omission of the dispersion interaction) and to yield target data “appropriate for the condensed phase”.<sup>13</sup> Effectively, this method results in an overestimation of the gas phase interaction energies. This in turn facilitates development of useful non-polarizable force fields by ensuring that resulting gas phase dipole moments are overestimated,<sup>11</sup> implicitly accounting for the polarization induced by the aqueous environment.

Since interactions with water molecules form such an important part of the charge parameter optimization scheme within the CHARMM force fields, it is interesting to check how well the models obtained from the relative entropy optimization can reproduce heterodimeric interactions with water. For ACEM we have considered the seven water interaction geometries shown in Figure 6. The corresponding minimum interaction energies and distances are shown in Table 4. Given that the ADD-OPT model has been optimized based on liquid phase data, one might expect it to provide a poor representation of gas phase dimers. This indeed seems to be the case, with interaction energies too favorable by more than 1 kcal/mol on average. Interestingly, however, in terms of the interaction energies, both the average difference and RMSD obtained with the ADD-OPT model are within 0.05 kcal/mol of the equivalent values obtained with the CGenFF force field, which was explicitly optimized based on the reproduction of interactions with water molecules.<sup>13</sup> This very good agreement suggests that the use of target quantum mechanical data which overestimates the interaction energies approximately accounts for the difference between the gas phase and aqueous environments by requiring the individual molecules to be overpolarized in the gas phase.

**1.3 Cyclohexane**—Water and acetamide are both important in the simulation of biomolecules, but neither molecule possesses any significant conformational flexibility. One of the objectives of this work is to use the newly-optimized additive force fields to enhance sampling in the polarizable simulations by connecting the different models via a replica exchange (“resolution exchange”) scheme. In order to test this approach, it is useful to study a system with high barriers, in which there is an equilibrium that is reached only relatively slowly. A natural example of such an equilibrium is the conformational exchange within the cyclohexane molecule (Figure 1). Cyclohexane can occupy several distinct conformations,

most notably the “chair”, “boat” and “twist-boat” conformations. The chair conformation is the lowest in energy, separated from the twist-boat minimum, which is 6 kcal/mol higher in energy, by a barrier of around 12 kcal/mol. As such, a typical simulation at room temperature would be expected to interconvert between chair and non-chair conformations only infrequently. Replica exchange simulations could be a route to enhancing sampling of this conformational interconversion. An alternative example could be a polypeptide conformational equilibrium; however at the time of writing, the protein parameters for the CHARMM Drude model were not yet available.

With only two distinct atom types, cyclohexane has only one free charge parameter that can be optimized and is well within the capabilities of the gradient-based optimization scheme described in Methods. The optimized charge parameters are shown in Table 1, and with these in place, a good reproduction of thermodynamic (Table S7) and structural (Figure S3) properties is obtained. Since we are interested in assessing the ability of the new force field to accelerate sampling of conformational interconversions we have also calculated the gas phase potential energy surface connecting the chair, boat and twist-boat conformations (Figure S4). We find that the ADD-OPT model provides a very good reproduction of the potential energy surface calculated with the polarizable model.

## 2. Replica Exchange Simulations

One of the principal motivations for optimizing non-polarizable analogues of polarizable models is that, by achieving the best possible overlap between the configurational distributions obtained with the two models, we maximize the probability of being able to perform replica exchange between the two models. Because the non-polarizable model is more computationally efficient, performing this replica exchange should allow us to enhance sampling in the polarizable force field. As a first test of this approach, we considered liquid water. The criterion for determining whether exchange will occur comes from considering the energy difference distributions shown in Equation 7. Initial tests showed that there was little overlap in these distributions obtained from simulations performed with the polarizable and nonpolarizable force fields, so that it was not possible to exchange directly between them (Figure S5). However, we also note that there was much better overlap between SWM4-NDP and SWM4A-OPT model than between SWM4-NDP and either of the TIP3P or TIP4P-Ew models (Figure S5), demonstrating that a specifically matched force field will facilitate resolution exchange.

By employing intermediate systems in which the polarizability was gradually reduced, we were able to couple the SWM4-NDP and SWM4A-OPT force fields. Specifically, results from a scheme having five distinct replicas are shown in Figure 7. In this simulation, replica 1 used the full SWM4-NDP water model, and replica 5 used the SWM4A-OPT model. The remaining replicas were then constructed as linear weighted averages of these two models: the polarizabilities of replicas 2, 3 and 4 were respectively 75%, 50% and 25% of that of SWM4-NDP. For each set of polarizabilities, the fixed charges were re-optimized against the fully polarizable target model, using the relative entropy optimization described above. By using such a scheme, we were able to obtain overlap between the energy difference distributions (Equation 7) obtained from adjacent replicas (Figure 7a), and hence exchange between them. Figure 7b shows that this scheme allows the simulations to move through all of the available replicas.

The results from the replica exchange simulations of water are important in their own right: they show that replica exchange between polarizable and non-polarizable force fields is possible, and since water is the largest component of any biomolecular simulation, this is a key requirement. However, since this water model is a simple molecule with no internal degrees of freedom, it is not possible to assess the impact of resolution exchange on

sampling efficiency. To test this more thoroughly, we therefore also considered liquid cyclohexane. Cyclohexane can adopt a number of distinct conformations. As noted above, the global energy minimum is the chair conformation, with the distorted boat, or “twist-boat” minimum approximately 6 kcal/mol higher in energy and these two conformations separated by an energetic barrier of around 12 kcal/mol. Individual twist-boat conformations are separated by boat transition states at an energy of around 7.5 kcal/mol. Taking these energy differences into account, one would expect that, at equilibrium, more than 99% of all cyclohexane molecules would be in the chair conformation. To test the sampling efficiency of various simulation protocols we constructed a box of 125 cyclohexane molecules in which every molecule was in the boat conformation. After an initial 100 ps equilibration, during which 15 molecules converted to the chair conformation, we adopted three different protocols for sampling the cyclohexane system. The first was a standard equilibrium simulation using the polarizable force field at 300 K; the second was a temperature REMD simulation that included eight replicas, all simulated using the CHARMM Drude polarizable force field, at temperatures of 300, 310, 325, 340, 355, 370, 385 and 400 K; and the third was a combined resolution and temperature replica exchange simulation. This also had 8 replicas: the first used the polarizable force field at 300 K; the second used the ADD-OPT model at 310 K, and the remainder all used the ADD-OPT model at temperatures increasing in 15 K increments up to 400 K. Unlike the water simulations, no intermediate states were required to couple between the polarizable and non-polarizable force fields – direct exchanges occurred between the two; however, inclusion of more intermediates to bridge the polarizable and non-polarizable simulations should pose no problem, as we have demonstrated for water. To test the sampling achieved with the different schemes, we then measured how long it took for the cyclohexane box to reach a near-equilibrium point where fewer than 1% of molecules were not in a chair conformation. The results of these simulations are shown in Figure 8, and clearly indicate a modest enhancement of sampling by the use of the resolution exchange scheme. With the resolution exchange, equilibrium is achieved with around 400 ps of simulation in the low temperature, polarizable replica. With the polarizable temperature replica exchange scheme, it takes more than 800 ps of simulation with the low temperature replica before equilibrium is reached. We are achieving an approximately 2-fold speed-up by using the resolution exchange scheme. One of the main reasons for this is that the simulation timestep we can use is determined by the fastest motions present in the system. In the Drude polarizable force field, the fastest motions are those of the Drude particles, which dictate that a timestep of 1 fs is required. In the non-polarizable force field, a timestep of 2 fs can be used. This means that, in the combined resolution/temperature exchange scheme, in the time that the polarizable replica takes to complete 100 ps of simulation, each other replica has completed 200 ps of simulation. In comparison, in the polarizable temperature replica exchange simulations, each replica will have completed only 100 ps. The degree of acceleration is expected to be larger for polarizable models with a higher computational cost.

## Concluding Remarks

We have developed a well-defined protocol for determining point charges in additive force fields, which has two advantages over standard methods for deriving such charges: (i) it provides an algorithmic method for including average polarization effects by explicitly matching the distribution in the condensed phase obtained with a polarizable force field, rather than by *ad hoc* procedures, and (ii) it permits the additive force field to be used to enhance sampling in the polarizable force field via multiscale simulations. The procedure is of course not limited to the CHARMM polarizable force field and could be generalized to target either more computationally demanding force fields, or even Born-Oppenheimer molecular dynamics simulations.<sup>25</sup>

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

CMB is supported by a FEBS Return to Europe Fellowship. RBB was supported by a Royal Society University Research Fellowship and by the Intramural research program of the National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health (NIH). The authors benefitted from a HECToR Class 2a computing allocation.

## Abbreviations

**RESP**      Restrained Electrostatic Potential fitting

## References

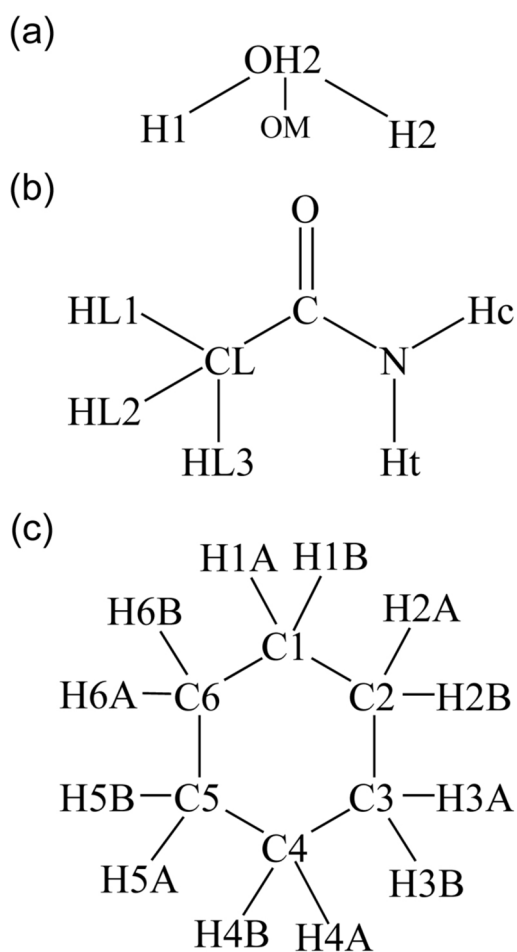
1. Karplus M, McCammon JA. *Nat Struct Biol.* 2002; 9:646–652. [PubMed: 12198485]
2. Baker, CM.; Darian, E.; MacKerell, AD, Jr. *Innovation in Biomolecular Modeling and Simulations.* Schlick, T., editor. Vol. 1. RSC Publications; Cambridge, UK: 2012. p. 23-50.
3. Ponder JW, Wu C, Ren P, Pande VS, Chodera JD, Schnieders MJ, Haque I, Mobley DL, Lambrecht DS, DiStasio RA Jr, Head-Gordon M, Clark GNI, Johnson ME, Head-Gordon T. *J Phys Chem B.* 2010; 114:2549–2564. [PubMed: 20136072]
4. Momany FA. *J Phys Chem.* 1978; 82:592–601.
5. Williams DE. *Biopolymers.* 1990; 29:1367–1386.
6. Reynolds CA, Essex JW, Richards WG. *Chem Phys Lett.* 1992; 199:257–260.
7. Stouch TR, Williams DE. *J Comput Chem.* 1992; 13:622–632.
8. Cieplak P, Cornell WD, Bayly C, Kollman PA. *J Comput Chem.* 1995; 16:1357–1377.
9. Reynolds CA, Essex JW, Richards WG. *J Am Chem Soc.* 1992; 114:9075–9079.
10. Duan Y, Wu C, Chowdhury S, Lee MC, Xiong G, Zhang W, Yang R, Cieplak P, Luo R, Lee T, Caldwell J, Wang J, Kollman P. *J Comput Chem.* 2003; 24:1999–2012. [PubMed: 14531054]
11. MacKerell AD Jr, Bashford D, Bellot M, Dunbrack RL Jr, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE III, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M. *J Phys Chem B.* 1998; 102:3586–3616.
12. Foloppe N, MacKerell ADJ. *J Comput Chem.* 2000; 21:86–104.
13. Vannomeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, Darian E, Guvench O, Lopes P, Vorobyov I, MacKerell AD. *J Comput Chem.* 2009; 31:671–690.
14. Cornell WD, Cieplak P, Bayly CI, Kollman P. *J Am Chem Soc.* 1993; 115:9620–9631.
15. Cerutti DS, Rice JE, Swope W, Case DA. *J Phys Chem B.* 2013
16. Kumar R, Skinner JL. *J Phys Chem B.* 2008; 112:8311–8318. [PubMed: 18570461]
17. Tainter CJ, Pieniazek PA, Lin YS, Skinner JL. *J Chem Phys.* 2011; 134:184501. [PubMed: 21568515]
18. Lopes PEM, Roux B, MacKerell AD Jr. *Theor Chem Acc.* 2009; 124:11–28. [PubMed: 20577578]
19. Cieplak P, Dupradeau FY, Duan Y, Wong J. *J Phys Condens Matter.* 2009; 21:333102. [PubMed: 21828594]
20. Jiang W, Hardy DJ, Phillips JC, Mackerell AD Jr, Schulten K, Roux B. *J Phys Chem Lett.* 2011; 2:87–92. [PubMed: 21572567]
21. Cieplak P, Caldwell J, Kollman P. *J Comput Chem.* 2001; 22:1048–1057.
22. Ren P, Ponder JW. *J Comput Chem.* 2002; 23:1497–1506. [PubMed: 12395419]
23. Lindorff-Larsen K, Piana S, Dror RO, Shaw DE. *Science.* 2011; 334:517–520. [PubMed: 22034434]

24. Best RB. *Curr Opin Struct Biol.* 2012; 22:52–61. [PubMed: 22257762]
25. Izvekov S, Parrinello M, Burnham CJ, Voth GA. *J Chem Phys.* 2004; 120:10896. [PubMed: 15268120]
26. Iuchi S, Izvekov S, Voth GA. *J Chem Phys.* 2007; 126:124505. [PubMed: 17411142]
27. Lyman E, Ytreberg FM, Zuckerman DM. *Phys Rev Lett.* 2006; 96:028105. [PubMed: 16486650]
28. Shell MS. *J Chem Phys.* 2008; 129:144108. [PubMed: 19045135]
29. Chaimovich A, Shell MS. *Phys Chem Chem Phys.* 2009; 11:1901–1915. [PubMed: 19280001]
30. Chaimovich A, Shell MS. *Phys Rev E.* 2010; 81:060104.
31. Chaimovich A, Shell MS. *J Chem Phys.* 2011; 134:094112. [PubMed: 21384955]
32. Shell MS. *J Chem Phys.* 2012; 137:084503. [PubMed: 22938246]
33. Kullback S, Leibler RA. *Ann Math Stat.* 1951; 22:79–86.
34. Carmichael SP, Shell MS. *J Phys Chem B.* 2012; 116:8383–8393. [PubMed: 22300263]
35. Wang LP, Chen J, Van Voorhis T. *J Chem Theory Comput.* 2013; 9:452–460.
36. Brooks BR, Brooks CL III, Mackerell AD Jr, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodosek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodstock HL, Wu X, Yang W, York DM, Karplus M. *J Comput Chem.* 2009; 30:1545–1614. [PubMed: 19444816]
37. Lamoureux G, Harder E, Vorobyov IV, Roux B, MacKerell ADJ. *Chem Phys Lett.* 2006; 418:245–249.
38. Harder E, Anisimov VM, Vorobyov IV, Lopes PEM, Noskov SY, Mackerell AD Jr, Roux B. *J Chem Theory Comput.* 2006; 2:1587–1597.
39. Vorobyov IV, Anisimov VM, Greene S, Venable RM, Moser A, Pastor RW, Mackerell AD Jr. *J Chem Theory Comput.* 2007; 3:1120–1133.
40. Swope WC, Anderson HC, Berens PHRWK. *J Chem Phys.* 1982; 76:637–649.
41. Lamoureux G, Roux B. *J Chem Phys.* 2003; 119:3025–3039.
42. Nose S. *Mol Phys.* 1984; 52:255–268.
43. Hoover WG. *Phys Rev A.* 1985; 31:1695–1697. [PubMed: 9895674]
44. Martyna GJ, Tobias DJ, Klein ML. *J Chem Phys.* 1994; 101:4177–4189.
45. Ryckaert JP, Ciccotti G, Berendsen HJ. *J Comput Phys.* 1997; 23:327–341.
46. Lague P, Pastor RW, Brooks BR. *J Phys Chem B.* 2004; 108:363–368.
47. Darden T, York D, Pedersen L. *J Chem Phys.* 1993; 98:10089–10092.
48. Allen, MP.; Tildesley, DJ. *Computer Simulation of Liquids.* 1. Clarendon; Oxford: 1987.
49. Tu Y, Laaksonen A. *Chem Phys Lett.* 2000; 329:283–288.
50. Kollman P. *Chem Rev.* 1993; 93:2395–2417.
51. Deng Y, Roux B. *J Phys Chem B.* 2004; 108:16567–16576.
52. Baker CM, Lopes PEM, Zhu X, Roux B, Mackerell AD Jr. *J Chem Theory Comput.* 2010; 6:1181–1198. [PubMed: 20401166]
53. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. *J Chem Phys.* 1983; 79:926–935.
54. Horn HW, Swope WC, Pitner JW, Madura JD, Dick TJ, Hura GL, Head-Gordon T. *J Chem Phys.* 2004; 120:9665–9678. [PubMed: 15267980]
55. Sugita Y, Okamoto Y. *Chem Phys Lett.* 1999; 314:141–151.
56. Lamoureux G, Mackerell AD Jr, Roux B. *J Chem Phys.* 2003; 119:5185–5197.
57. Anisimov VM, Lamoureux G, Vorobyov IV, Huang N, Roux B, Mackerell AD Jr. *J Chem Theory Comput.* 2005; 1:153–168.
58. Vorobyov IV, Anisimov VM, Mackerell AD Jr. *J Phys Chem B.* 2005; 109:18988–18999. [PubMed: 16853445]
59. Lopes PEM, Lamoureux G, Roux B, Mackerell AD Jr. *J Phys Chem B.* 2007; 111:2873–2885. [PubMed: 17388420]

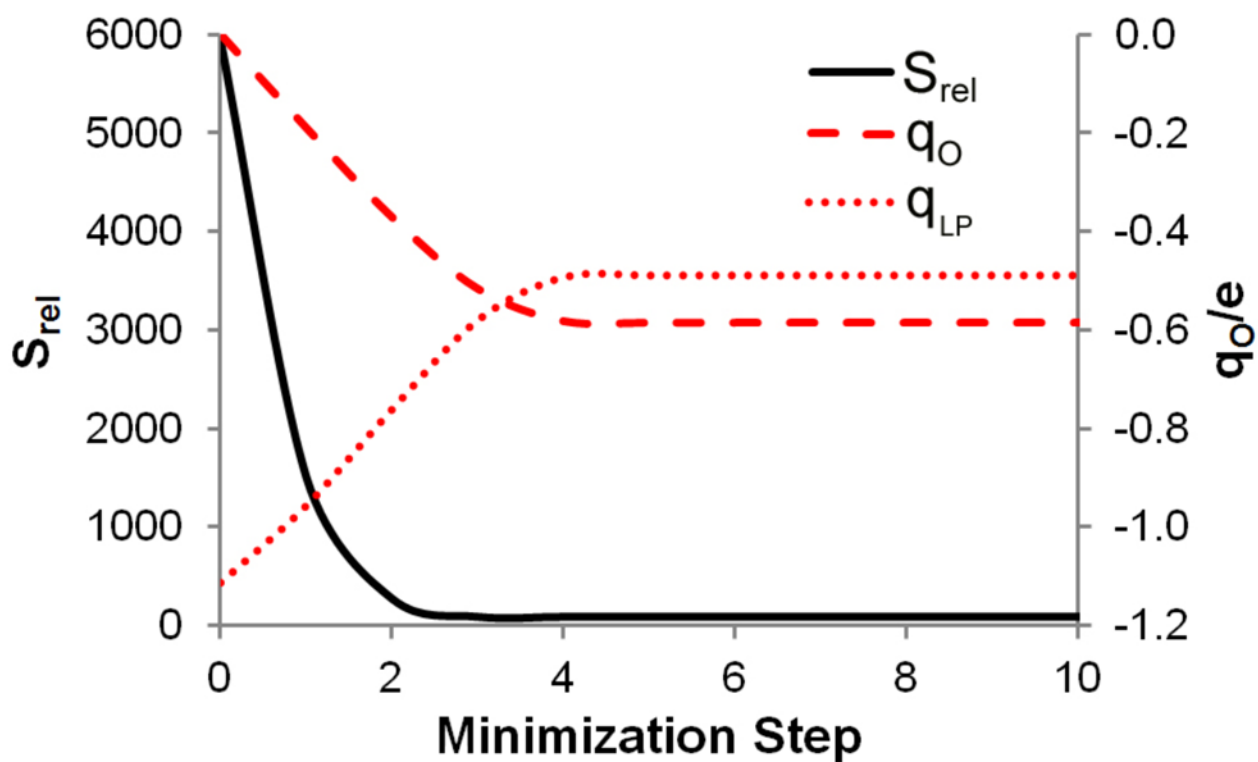
60. Anisimov VM, Vorobyov IV, Roux B, Mackerell AD Jr. *J Chem Theory Comput.* 2007; 3:1927–1946. [PubMed: 18802495]
61. Harder E, Anisimov VM, Whitfield TW, Mackerell AD Jr, Roux B. *J Phys Chem B.* 2008; 112:3509–3521. [PubMed: 18302362]
62. Lopes PEM, Lamoureux G, Mackerell AD Jr. *J Comput Chem.* 2009; 30:1821–1838. [PubMed: 19090564]
63. Baker CM, Mackerell AD Jr. *J Mol Model.* 2010; 16:567–576. [PubMed: 19705172]
64. Yu H, Whitfield TW, Harder E, Lamoureux G, Vorobyov I, Anisimov VM, Mackerell AD Jr. *J Chem Theory Comput.* 2010; 6:774–786. [PubMed: 20300554]
65. Zhu X, Mackerell AD Jr. *J Comput Chem.* 2010; 31:2330–2341. [PubMed: 20575015]
66. Baker CM, Anisimov VM, Mackerell AD Jr. *J Phys Chem B.* 2011; 115:580–596. [PubMed: 21166469]
67. Orabi EA, Lamoureux G. *J Chem Theory Comput.* 2012; 8:182–193.
68. Yu W, Lopes PEM, Roux B, MacKerell ADJ. *J Chem Phys.* 2013; 138:034508. [PubMed: 23343286]
69. Macleod NA, Butz P, Simons JP, Grant GH, Baker CM, Tranter GE. *Isr J Chem.* 2004; 144:27–36.
70. Barciszewski J, Jurczak J, Porowski S, Specht T, Erdmann VA. *Eur J Biochem.* 1999; 260:293–307. [PubMed: 10095763]
71. Macleod NA, Butz P, Simons JP, Grant GH, Baker CM, Tranter GE. *Phys Chem Chem Phys.* 2005; 7:1432–1440. [PubMed: 19787965]
72. Baker CM, Grant GH. *Chem Commun.* 2006:1387–1389.
73. Levy Y, Onuchic JN. *Annu Rev Biophys Biomol Struct.* 2006; 35:389–415. [PubMed: 16689642]
74. Baker CM, Grant GH. *J Phys Chem B.* 2007; 111:9940–9954. [PubMed: 17672488]
75. Huggins DJ, Tidor B. *Prot Eng Des Sel.* 2011; 24:777–789.
76. Mark P, Nilsson L. *J Phys Chem A.* 2001; 2001:9954–9960.
77. Huggins DJ. *J Chem Phys.* 2012; 136:064518. [PubMed: 22360206]
78. Guillot B. *J Mol Liq.* 2002; 2002:219–260.
79. Gladich I, Roeselova M. *Phys Chem Chem Phys.* 2012; 14:11371–11385. [PubMed: 22801804]
80. Wang J, Cieplak P, Cai Q, Hsieh MJ, Wang J, Duan Y, Luo R. *J Phys Chem B.* 2012; 116:7999–8008. [PubMed: 22712654]
81. Soper AK, Phillips MG. *Chem Phys.* 1986; 107:47–60.
82. Berendsen HJC, Grigera JR, Straatsma TP. *J Phys Chem.* 1987; 91:6269–6271.
83. Vinter JG. *J Comput-Aided Mol Des.* 1994; 8:653–668. [PubMed: 7738602]
84. Dixon RW, Kollman PA. *J Comput Chem.* 1997; 18:1632–1646.
85. Baker CM, Grant GH. *J Chem Theory Comput.* 2006; 2:947–955.
86. Baker CM, Grant GH. *J Chem Theory Comput.* 2007; 3:530–548.
87. Kirkpatrick S, Gelatt CD Jr, Vecchi MP. *Science.* 1983; 220:671–680. [PubMed: 17813860]
88. Guvench O, MacKerell AD Jr. *J Mol Model.* 2008; 14:667–679. [PubMed: 18458967]
89. MacKerell AD Jr, Bashford D, Bellot M, Dunbrack JRL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE III, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Kuczera J, Yin D, Karplus M. *J Phys Chem B.* 2000; 102:3586–3616.
90. Best RB, De Sancho D, Mittal J. *Biophys J.* 2012; 102:1462–1467. [PubMed: 22455930]
91. Dyke TR, Muentner JS. *J Chem Phys.* 1973; 59:3125–3127.
92. Dyke TR, Mack KM, Muentner JS. *J Chem Phys.* 1977; 66:498–510.
93. Chase MW Jr. *J Phys Chem Ref Data, Monograph.* 1998; 9:1–1951.
94. Krynicki K, Green CD, Sawyer DW. *Faraday Discuss.* 1978; 66:199–208.
95. Fernandez DP, Mulev Y, Goodwin ARHL, SJMH. *J Phys Chem Ref Data.* 1995; 24:33–69.
96. Ben-Naim A, Marcus Y. *J Chem Phys.* 1984; 81:2016–2027.
97. Pearson RG. *J Am Chem Soc.* 1986; 108:6109–6114.

98. Hermans J, Pathiaseril A, Anderson A. *J Am Chem Soc.* 1988; 110:5982–5986. [PubMed: 22148769]
99. Abraham MH, Whiting GS, Fuchs R, Chambers EJ. *J Chem Soc Perk T.* 1990; 2:291–300.
100. Vargas R, Garza J, Friesner RA, Stern H, Hay BP, Dixon DA. *J Phys Chem A.* 2001; 105:4963–4968.
101. Kelly CP, Cramer CJ, Truhlar D. *J Chem Theory Comput.* 2005; 1:1133–1152.

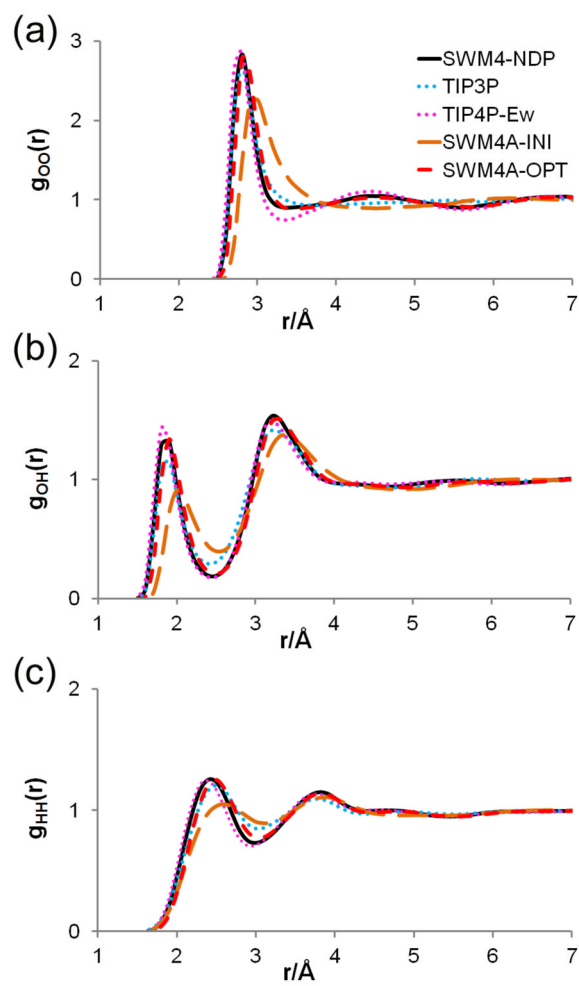




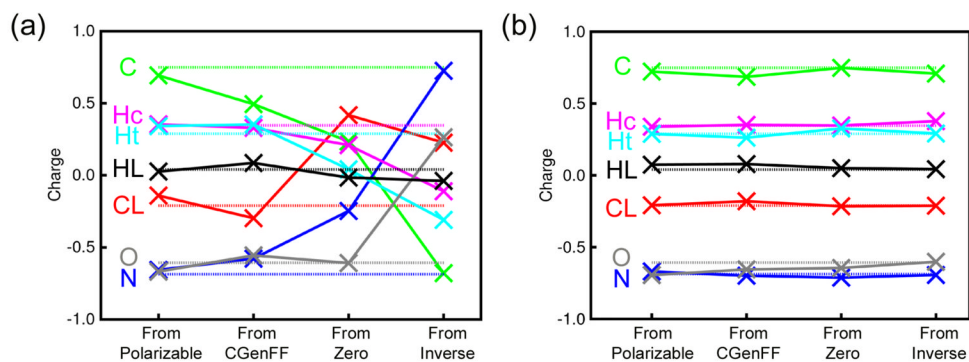
**Figure 1.**  
Atom naming used in this work for (a) water (b) acetamide and (c) cyclohexane.



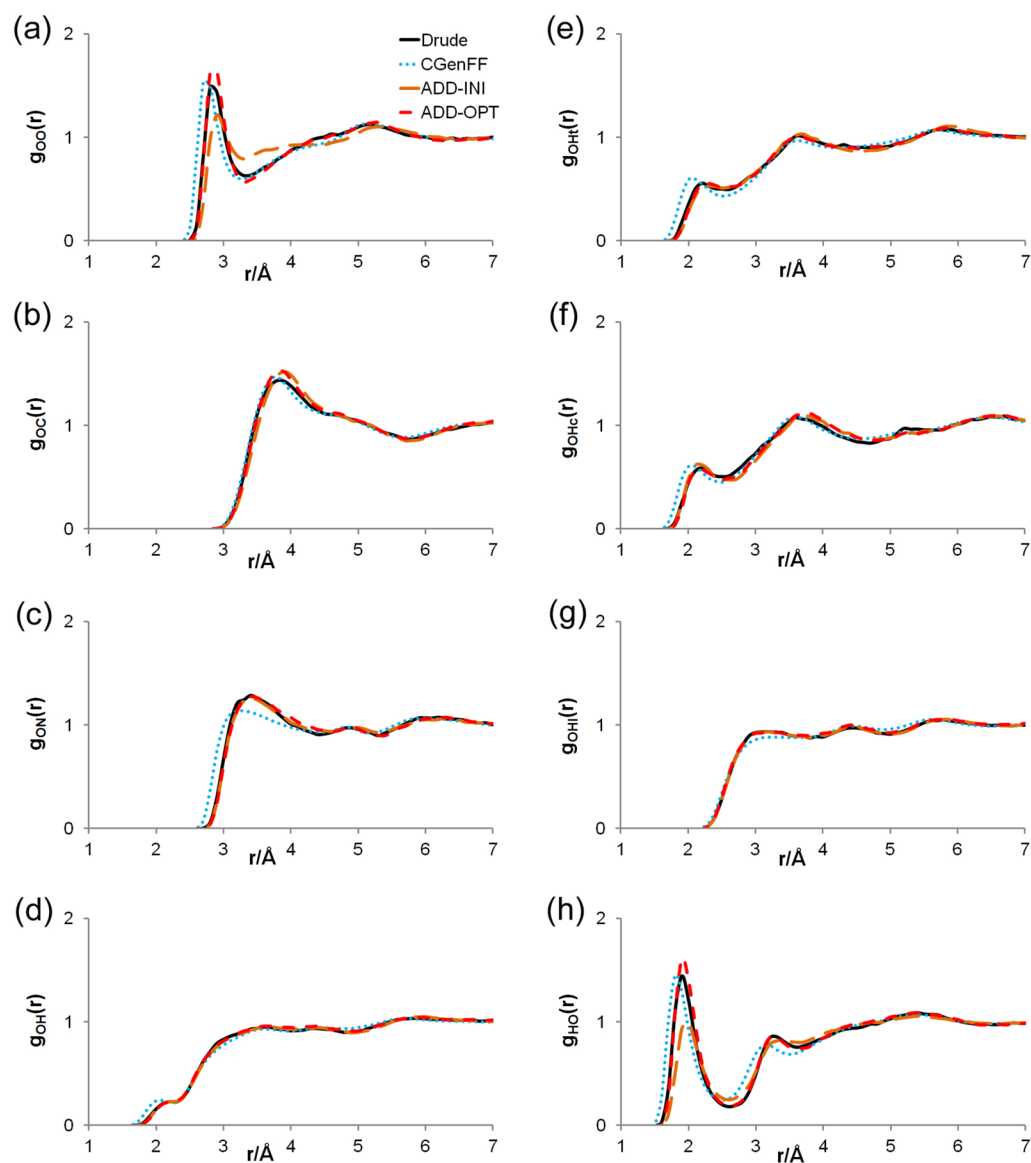
**Figure 2.** Water model optimization. Variation of the charge on the oxygen and dummy atoms,  $q_O$  and  $q_{LP}$ , respectively, during Newton-Raphson optimization of the relative entropy,  $S_{rel}$ . The starting point for the minimization is the SWM4A-INI model, and the end point is the SWM4A-OPT model.



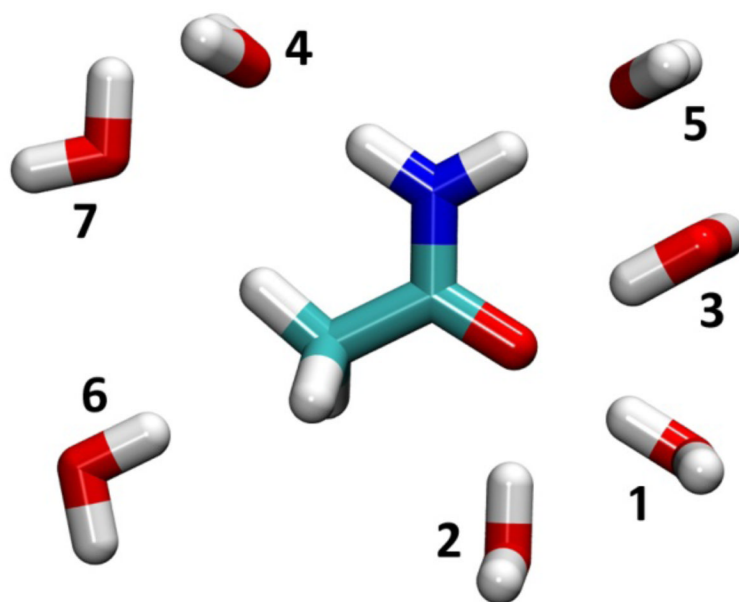
**Figure 3.** Radial distribution functions for simulations of liquid water with various force fields (see text for details).



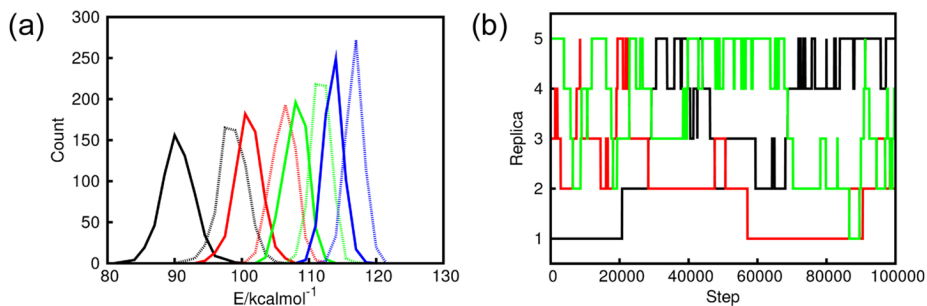
**Figure 4.** Graphs showing the final acetamide charges obtained from relative entropy optimizations starting from four different sets of initial acetamide charges. (a) shows charges resulting from calculations in which the relative entropy is optimized using minimization without restraints, and (b) shows charges resulting from calculations in which the relative entropy is optimized using a Monte Carlo simulated annealing algorithm including a restraint potential as shown in Equation 9 with a force constant  $k_q = 200$ . In both panels, the charges used in the polarizable model are indicated by the dotted lines.



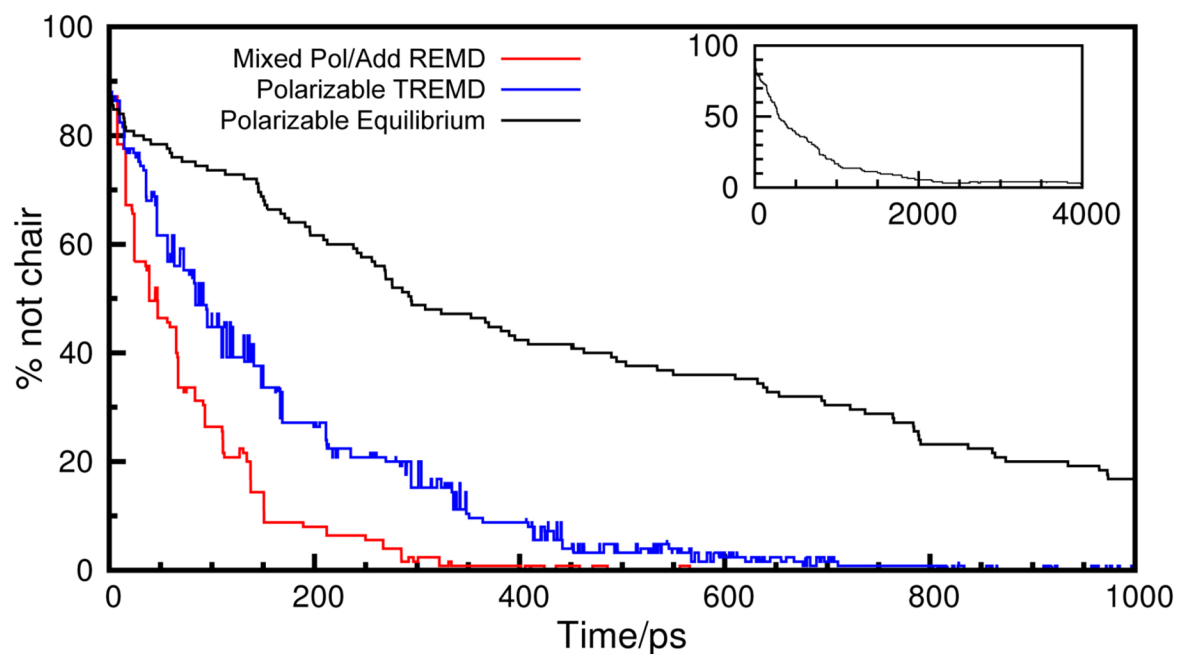
**Figure 5.** Radial distribution functions obtained from simulations of acetamide in water, calculated using various force fields. Radial distribution functions are described as  $g_{ab}(r)$ , where  $a$  represents the water atom type, and  $b$  represents the acetamide atom type.



**Figure 6.** Orientations of interactions between acetamide and water. Analysis was performed on monohydrates; for the purpose of visualization all water interaction orientations are displayed simultaneously in this figure.



**Figure 7.** Hamiltonian REMD simulations of liquid water performed using five distinct replicas. Replica 1 corresponds to the SWM4-NDP model and replica 5 to the SWM4A-OPT model. Intermediate replicas employ weighted averages of these two models: replica 2 is 75% SWM4-NDP; replica 3 is 50% SWM4-NDP, and replica 4 is 25% SWM4-NDP. (a) Overlap of energy difference distributions from adjacent replicas, calculated as described for Equation 7. Matching colors are used to compare overlap between distributions from adjacent replicas; for each pair of replicas, solid lines represent  $\Delta_i$ , with dashed lines representing  $\Delta_{i+1}$ . (b) Tracking the progress of three of the replicas over the course of a 100 ps MD simulation.



**Figure 8.**

Convergence toward equilibrium for a box of 125 cyclohexane molecules using several simulation protocols. Simulations were started from a configuration in which 88% of molecules were in the twist-boat conformation, which is  $\sim 6$  kcal/mol higher in energy than the global energy minimum chair conformation. At each simulation step, we record the percentage of molecules that are not in the chair conformation: at equilibrium, one would expect  $< 1\%$  of molecules to be in anything other than a chair conformation. The inset panel shows the progress of the equilibrium simulation with the polarizable force field.



**Table 1**

Atomic charges used in the force fields described in the text. Atom naming is as shown in Figure 1.

	CHARMM Drude Model	Reference CHARMM Additive Model	Optimized Additive Model
Water			
	SWM4-NDP	TIP3P	SWM4A-OPT
OH2	0.00000	-0.83400	-0.58430
OM	-1.11466	N/A	-0.48830
H1/H2	0.55733	0.41700	0.53630
Acetamide			
	CHARMM Drude	CGenFF	ADD-OPT
CL	-0.21000	-0.27000	-0.20840
C	0.74900	0.55000	0.72210
N	-0.68700	-0.62000	-0.66910
Hc	0.34600	0.32000	0.33890
Ht	0.28900	0.30000	0.28870
O	-0.00100	-0.55000	-0.69450
HL1/HL2/HL3	0.04000	0.09000	0.07410
LPA	-0.37600	N/A	N/A
LPB	-0.23000	N/A	N/A
Cyclohexane			
	CHARMM Drude	CHARMM aliphatic	ADD-OPT
C	-0.08000	-0.18000	-0.06954
H	0.04000	0.09000	0.03477

Table 2

Water properties calculated using the polarizable and initial and optimized non-polarizable models. Experimental results and results obtained using the TIP3P water model are included for comparison.

	Experiment	SWM4-NDP <sup>g</sup>	TIP3P <sup>g</sup>	TIP4P-E <sub>w</sub> <sup>f</sup>	SWM4A-INI	SWM4A-OPT
Monomer						
Dipole (D)	1.85 <sup>c</sup>	1.85	2.35	2.32	1.85	2.46
Polarizability $\alpha$ ( $\text{\AA}^3$ )	1.44 <sup>d</sup>	0.98	N/A	N/A	N/A	N/A
Dimer						
$U_{\text{dimer}}$ (kcal/mol)	-5.40 <sup>e</sup>	-5.15	-6.50	-7.42	-4.57	-7.43
$d_{\text{OO}}$ ( $\text{\AA}$ )	2.98 <sup>e</sup>	2.83	2.74	2.72	2.88	2.80
$\theta_{\text{A}}$ (deg)	58 <sup>e</sup>	71	20	51	67	23
$H_{\text{dimer}}$ (D)	2.64 <sup>e</sup>	2.06	3.87	3.00	1.82	4.14
Bulk Liquid						
$\Delta H_{\text{vap}}$ (kcal/mol) <sup>a</sup>	10.51 <sup>f</sup>	10.50	10.44	11.72	7.38 $\pm$ 0.01	12.13 $\pm$ 0.01
$\Delta H_{\text{vap}}$ (kcal/mol) <sup>b</sup>	n/a	n/a	9.19	10.58	7.38 $\pm$ 0.01	10.27 $\pm$ 0.01
$V_{\text{m}}$ ( $\text{\AA}^3$ )	29.94 <sup>g</sup>	29.91	29.90	30.0	33.90 $\pm$ 0.01	30.77 $\pm$ 0.02
$\langle L \rangle$ (D)	2.95 <sup>h</sup>	2.46	2.35	2.32	1.85	2.46
D ( $10^{-5}$ cm <sup>2</sup> /s)	2.3 <sup>i</sup>	2.33 $\pm$ 0.02	5.1	2.4 $\pm$ 0.06	9.58 $\pm$ 0.02	2.81 $\pm$ 0.03
E	78 <sup>j</sup>	79 $\pm$ 3	92	63.9 $\pm$ 0.9	25.8 $\pm$ 0.2	76.1 $\pm$ 1.4
$\Delta G_{\text{hyd}}$ (kcal/mol) <sup>a</sup>	-4.4 - -6.3 <sup>k</sup>	-5.9 $\pm$ 0.1	-6.1	-7.09 $\pm$ 0.05	-4.08 $\pm$ 0.06	-7.56 $\pm$ 0.07
$\Delta G_{\text{hyd}}$ (kcal/mol) <sup>b</sup>	n/a	n/a	-4.85	-5.95 $\pm$ 0.05	-4.08 $\pm$ 0.06	-5.70 $\pm$ 0.07

<sup>a</sup>Excluding polarization correction;

<sup>b</sup>including polarization correction (Equation 8) with the experimental isotropic gas phase polarizability<sup>37</sup> of 1.44  $\text{\AA}^3$ .

<sup>c</sup>From Ref. 91.

<sup>d</sup>From Ref 37.

<sup>e</sup>From Ref. 92.

<sup>f</sup>From Ref. 93.

<sup>g</sup>From Ref. 37.

<sup>h</sup>From Ref. 68.

<sup>i</sup>From Ref. 94.

<sup>j</sup>From Ref. 95.

<sup>k</sup>From Refs. 96–99.

<sup>l</sup>From Ref. 54.

**Table 3**

Acetamide properties calculated using polarizable and initial and optimized additive models. Results obtained with the additive CGenFF model are included for comparison. Properties in aqueous solution are calculated using the SWM4-NDP water model with polarizable acetamide; the TIP3P water model with CGenFF acetamide, and the SWM4A-OPT water model with ADD-INI and ADD-OPT acetamide.

	Experiment	Drude	CGenFF	ADD-INI	ADD-OPT
Monomer					
Dipole (D)	4.0 <sup>c</sup>	3.7	4.2	4.1	4.9
Dimer					
$U_{\text{dimer}}$ (kcal/mol)	-14.35 <sup>d</sup>	-13.26	-10.98	-11.25	-13.17
$d_{\text{OHc}}$ (Å)	1.83 <sup>d</sup>	1.95	1.83	2.03	1.99
Aqueous Solution					
$\langle \mu \rangle$ (D)		4.6	4.2	4.1	4.9
$\Delta G_{\text{hyd}}$ (kcal/mol) <sup>a</sup>	-9.71 <sup>e</sup>	-9.18	-8.79	-7.45 ± 0.12	-11.34 ± 0.11
$\Delta G_{\text{hyd}}$ (kcal/mol) <sup>b</sup>	n/a	n/a	-8.49	-7.26 ± 0.12	-9.58 ± 0.11

<sup>a</sup>Excluding polarization correction;

<sup>b</sup>including polarization correction (in the absence of an experimental gas phase polarizability, we use the quantum mechanically calculated<sup>61</sup> trace of the molecular polarizability tensor, 5.9 Å<sup>3</sup>).

<sup>c</sup>From Ref. 61.

<sup>d</sup>From Ref. 100.

<sup>e</sup>From Ref. 101.

Table 4

Acetamide-water heterodimer interaction energies and distances.  $E_{\text{int}}$  is the energy at the minimum interaction distance for a given water orientation in kcal/mol;  $R_{\text{min}}$  is the minimum interaction distance for a given water orientation in Å. Water orientations are as shown in Figure 6. Calculations use the SWM4-NDP water model with polarizable acetamide; the TIP3P water model with CGenFF acetamide, and the SWM4A-OPT water model with ADD-  
INI and ADD-OPT acetamide.

	Drude		CGenFF		ADD- INI		ADD- OPT	
	$E_{\text{int}}$	$R_{\text{min}}$	$E_{\text{int}}$	$R_{\text{min}}$	$E_{\text{int}}$	$R_{\text{min}}$	$E_{\text{int}}$	$R_{\text{min}}$
Conf1	-5.33	1.93	-6.99	1.75	-6.45	1.91	-8.13	1.87
Conf2	-6.06	1.88	-6.68	1.77	-5.78	1.94	-7.39	1.89
Conf3	-7.14	1.82	-7.19	1.74	-6.27	1.90	-7.98	1.86
Conf4	-4.08	2.09	-6.55	1.89	-5.52	2.05	-5.76	2.05
Conf5	-3.95	2.04	-5.22	1.88	-5.10	2.01	-4.68	2.02
Conf6	-0.22	2.65	-0.68	2.60	-0.75	2.50	0.06	2.87
Conf7	-1.26	2.79	-2.74	2.66	-1.68	2.76	-1.96	2.67
Ave Diff			-1.14	-0.13	-0.50	-0.02	-1.11	0.00
RMSD			1.38	0.14	0.92	0.07	1.43	0.10