

# Polarizable Force Field for Peptides and Proteins Based on the Classical Drude Oscillator

Pedro E. M. Lopes,<sup>†</sup> Jing Huang,<sup>†</sup> Jihyun Shim,<sup>†</sup> Yun Luo,<sup>‡,§,||</sup> Hui Li,<sup>‡</sup> Benoît Roux,<sup>\*,‡</sup> and Alexander D. MacKerell, Jr.<sup>\*,†</sup>

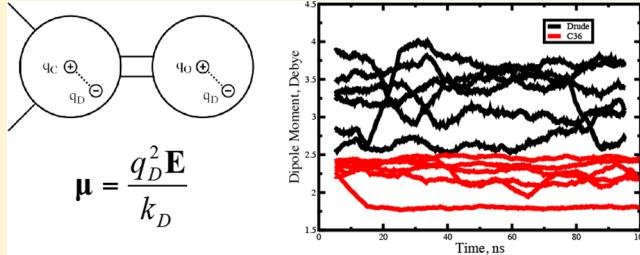
<sup>†</sup>Department of Pharmaceutical Sciences, University of Maryland School of Pharmacy, 20 Penn Street HSFII, Baltimore, Maryland 21201, United States

<sup>‡</sup>Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, Illinois 60637, United States

<sup>§</sup>Argonne Leadership Computing Facility, Argonne National Laboratory, 9700 South Cass Avenue, Building 240, Argonne, Illinois 60439, United States

## Supporting Information

**ABSTRACT:** Presented is a polarizable force field based on a classical Drude oscillator framework, currently implemented in the programs CHARMM and NAMD, for modeling and molecular dynamics (MD) simulation studies of peptides and proteins. Building upon parameters for model compounds representative of the functional groups in proteins, the development of the force field focused on the optimization of the parameters for the polypeptide backbone and the connectivity between the backbone and side chains. Optimization of the backbone electrostatic parameters targeted quantum mechanical conformational energies, interactions with water, molecular dipole moments and polarizabilities, and experimental condensed phase data for short polypeptides such as (Ala)<sub>5</sub>. Additional optimization of the backbone  $\phi$ ,  $\psi$  conformational preferences included adjustments of the tabulated two-dimensional spline function through the CMAP term. Validation of the model included simulations of a collection of peptides and proteins. This first generation polarizable model is shown to maintain the folded state of the studied systems on the 100 ns time scale in explicit solvent MD simulations. The Drude model typically yields larger RMS differences as compared to the additive CHARMM36 force field (C36) and shows additional flexibility as compared to the additive model. Comparison with NMR chemical shift data shows a small degradation of the polarizable model with respect to the additive, though the level of agreement may be considered satisfactory. However, the polarizable model shows improvement for the residues with significantly underestimated S<sup>2</sup> order parameters in the additive model. Analysis of dipole moments associated with the peptide backbone and tryptophan side chains show the Drude model to have considerably larger values than those present in C36, with the dipole moments of the peptide backbone enhanced to a greater extent in sheets versus helices and the dipoles of individual moieties observed to undergo large variations during the MD simulations. Although there are still some limitations, the presented model, termed Drude-2013, is anticipated to yield a molecular picture of peptide and protein structure and function that will be of increased physical validity and internal consistency in a computationally accessible fashion.



## INTRODUCTION

Empirical force field studies of peptides and proteins are widely used to understand the structural and dynamical properties of this biologically important class of molecules and relate them to their chemical functions. To date, these studies have largely been based on nonpolarizable, additive force fields, where the partial atomic charges of the system are fixed effective values accounting for induced electronic polarization in a mean-field manner, with the most widely used models including AMBER,<sup>1,2</sup> CHARMM,<sup>3–5</sup> GROMOS,<sup>6</sup> and OPLS,<sup>7,8</sup> among others.<sup>9,10</sup> Efforts to go beyond the additive approximation by incorporating an explicit treatment of electronic polarization have been ongoing for close to 30 years.<sup>11–13</sup> Already in 1976, Warshel and Levitt presented a polarizable model of lysozyme

in which polarization was incorporated via interacting induced point-dipoles.<sup>11</sup> Subsequent work over the following decades involved a range of developments that were critical to allow computationally efficient molecular dynamics (MD) simulations of solvated biological macromolecules based on polarizable models.<sup>14–18</sup> Such models, in which the electronic properties vary as a function of environment, are anticipated to yield a more physically realistic and consistent model, which would hopefully be more capable of reproducing a wide range of experimentally quantifiable observables accurately.

Received: September 2, 2013

Published: October 17, 2013

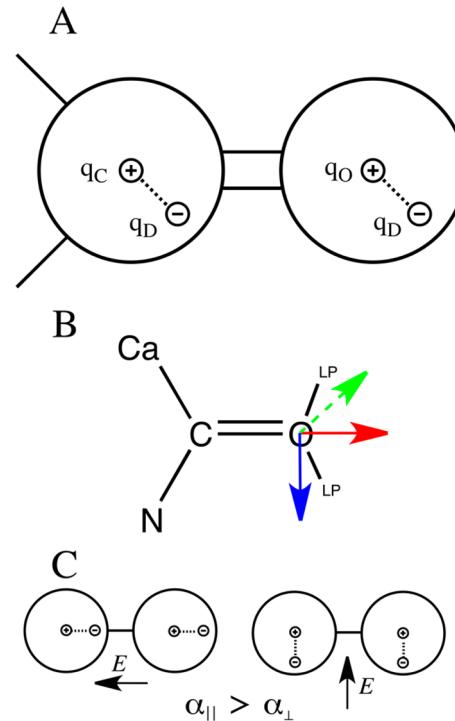


Toward the goal of a polarizable protein force field (FF) amenable to MD simulations, Berne, Friesner and co-workers introduced both induced dipole and fluctuating charge polarizable models,<sup>19</sup> as well as combinations thereof,<sup>20,21</sup> reporting gas phase protein simulations in 2002,<sup>22</sup> followed by a simulation in explicit solvent in 2005.<sup>23,24</sup> Patel and Brooks presented explicit solvent simulations of proteins in 2004 using a polarizable model based on a dynamically fluctuating charge model.<sup>25,26</sup> AMOEBA, which models the molecular charge distribution using a multipole representation along with induced point dipoles, was originally reported in 2002<sup>27</sup> and has been applied for studies of ligand binding to proteins.<sup>28–32</sup> While polarizable MD simulations using these, as well as other models,<sup>33,34</sup> have been presented, they are typically short in duration, being a few nanoseconds or less, such that the models have yet to be rigorously evaluated as well as used in application studies. A recent exception is a study on lysozyme using the fluctuating charge model implemented in CHARMM,<sup>35</sup> where simulations of up to 14 ns displayed RMS differences larger than those from the additive CHARMM22 force field.<sup>3</sup> More recently, an updated AMOEBA force field for proteins has been presented.<sup>36</sup> The model was optimized targeting both quantum mechanical and experimental data on model compounds and small peptides and used for simulations of a collection of small proteins for 30 ns. Thus, in practice it has clearly been a challenge to bridge the gap from the construction of a physically motivated polarizable force field for proteins at the conceptual level, and its implementation into a computationally efficient form at the practical level in order to allow one to undertake MD simulation studies for a range of systems and evaluate the robustness and accuracy of the model.

Development of the Drude polarizable force field in our laboratories has been ongoing since 2000.<sup>37</sup> Those efforts have led to the development of water models<sup>37–39</sup> and parameters for a collection of small molecules representative of the functional groups in proteins, nucleic acids, lipids, and carbohydrates,<sup>40–48</sup> as well as for ions.<sup>49,50</sup> In addition, MD simulation studies of a DNA duplex in solution with counterions<sup>51</sup> and of a dipalmitoylphosphatidylcholine (DPPC) bilayer and monolayer were reported.<sup>52</sup> While those studies were based on early models aimed at illustrating specific aspects of induced polarization, they were not based on optimized force field parameters suitable for a wide range of systems. More recently, we have completed refined models for DPPC<sup>53</sup> and acyclic polyalcohols.<sup>54</sup> Thus, progress is being made in extending the Drude polarizable force field from small molecules to biologically relevant macromolecular systems. In the present work, we extend those efforts to a polypeptide force field applicable for MD simulation studies of peptides and proteins. Emphasis is placed on optimization of the polypeptide backbone parameters and the connectivity between the backbone and the side chains, whose parameters were obtained from our previous publications with additional adjustments as outlined below. The final model, which will be named Drude-2013, but is referred to as Drude-3 in the present manuscript, is then applied to MD simulations of a collection of peptides and proteins to assess the validity and overall accuracy of the model. The results show that the model yields structural properties, including sampling of backbone and side chain conformations in agreement with experimental results at a level similar to that obtained with state-of-the-art additive force fields. We note that the range of proteins studied and the time scales of the present simulations are not adequate to fully probe the ability of the

model to reproduce a wide range of experimental observables as has been done over time with the additive models,<sup>55–63</sup> such that future improvements in the model are anticipated.

**Classical Drude Oscillator Force Field.** Explicit polarization in the classical Drude oscillator implementation (also referred to as a Shell or Charge-On-Spring (COS) model<sup>64,65</sup>) is based on attaching a charged auxiliary particle with a harmonic spring to the nucleus or atomic core of the parent atom (Figure 1a). The classical Drude originated from the



**Figure 1.** Example of the Drude oscillator model in the context of a carbonyl ( $C=O$ ) group. (A) The carbon and oxygen atoms shown as spheres with the atomic core (or nucleus) and the Drude particle, with the associated partial atomic charges for the nucleic,  $q_C$  and  $q_O$ , for carbon and oxygen, respectively, and  $q_D$  for the Drude particles, where the value of  $q_D$  varies for the different atom types. (B) Carbonyl group in the context of a peptide bond showing the approximate positions of the lone pairs and the tensor components  $A_{11}$  (red),  $A_{22}$  (green), and  $A_{33}$  (blue) that define the anisotropic atomic polarizability on the oxygen atom. (C) Schematic of the orientation of the Drude particles in a diatomic relative to the atomic centers with the electric field,  $E$ , parallel and perpendicular to the covalent bond.

quantum mechanical (QM) Drude model used to treat dispersion<sup>66,67</sup> and has been subsequently used for ionic crystals, simple liquids, water, and ions and in QM/MM calculations.<sup>17</sup> In the classical Drude model, the induced atomic dipole,  $\mu$ , in the presence of an electric field,  $E$ , is

$$\mu = \frac{q_D^2 E}{k_D} \quad (1)$$

such that the atomic polarizability,  $\alpha$ , is equivalent to the Drude charge,  $q_D$ , divided by the force constant on the harmonic spring,  $k_D$ ,

$$\alpha = \frac{q_D^2}{k_D} \quad (2)$$

In practice, the unperturbed static partial atomic charge of the atom,  $q$ , is distributed between the parent atom and the virtual Drude particle.

For the sake of simplicity, the current implementation of the classical Drude model introduces atomic polarizabilities only to non-hydrogen atoms. However, this is adequate to accurately reproduce molecular polarizabilities, as seen in a number of published studies.<sup>41,42,47</sup> In the case of hydrogen bond acceptors, it has been shown that inclusion of anisotropic atomic polarizability improves nonbond interactions involving acceptors as a function of orientation, especially in the case of interactions with ions.<sup>44</sup> Anisotropic atomic polarization is defined as the isotropic polarizability times a tensor  $A$ , with trace = 3 (Figure 1b).  $A$  is diagonal in a local reference frame defined by the vector,  $\mathbf{A}11$ , between the atom on which anisotropic polarizability is being assigned and a covalently bound atom or, more generally, any particle, such as a lone pair. The third and fourth atoms, or particles, define the  $\mathbf{A}22$  vector, with the  $\mathbf{A}33$  vector being orthogonal to  $\mathbf{A}11$  and  $\mathbf{A}22$ . As the trace of  $A$  is set to 3, only the values of  $\mathbf{A}11$  and  $\mathbf{A}22$  need to be set, such that  $\mathbf{A}33 = 3 - \mathbf{A}11 - \mathbf{A}22$ . For example, the specification in the residue-topology-file (RTF) within the CHARMM syntax in the case of the peptide backbone carbonyl O (Figure 1B), is “ANISOTROPY O C CA N A11 0.82322 A22 1.14332,” where the anisotropic polarizability is on the O, the  $\mathbf{A}11$  vector is along the C=O bond, and the  $\mathbf{A}22$  vector is along the CA–N direction. To further facilitate the ability of the model to treat nonbond interactions involving acceptors, virtual particles representative of lone pairs are included in the model.<sup>44</sup> In the case of the carbonyl O, this involves two LP sites (Figure 1b), whose positions were optimized to improve the interactions with water as a function of orientation.

An important feature in a polarizable force field is the ability of the induced atomic dipoles to interact with each other within a molecule to yield the correct molecular polarizability tensor.<sup>68–70</sup> Traditionally, in additive force fields, nonbond interactions between atoms covalently bound to each other (1,2 pairs) or the terminal atoms on a valence angle (1,3 pairs) are ignored while nonbond interactions between the terminal atoms of dihedral angles (1,4 pairs) and beyond are treated explicitly, with 1,4 pair interactions being scaled in some force fields.<sup>1,8</sup> However, with a polarizable force field, it is possible for all induced atomic dipoles to interact with each other, including 1,2 and 1,3 pairs, as a means to obtain the correct molecular polarization response. An example is the case of diatomic, as shown in Figure 1c. With the diatomic, when the electric field is parallel to the bond between the atoms, the atomic dipoles interact favorably, while when the field is perpendicular to the bond, the atomic dipoles interact unfavorably. It is these differences that yield a nontrivial anisotropic molecular polarizability tensor, which in the case of the diatomic will be larger parallel to the bond as compared to perpendicular to the bond. While it is useful to include these interactions in the polarizable force field, their spatial separation is such that the Coulombic approximation fails, disallowing the use of point charges for the 1,2 and 1,3 interactions. To overcome this, the electrostatic shielding treatment proposed by Thole is applied,<sup>69</sup> in which the coulomb interactions between charges  $i$  and  $j$  are modulated by a factor,  $S_{ij}$ , as shown in eq 3

$$S_{ij}(r_{ij}) = 1 - \left(1 + \frac{(a_i + a_j)r_{ij}}{2(\alpha_i\alpha_j)^{1/6}}\right) e^{-(a_i + a_j)r_{ij}/2(\alpha_i\alpha_j)^{1/6}} \quad (3)$$

where  $r_{ij}$  is the distance between the atoms,  $\alpha_i$  and  $\alpha_j$  are the respective atomic polarizabilities, and  $a_i$  and  $a_j$  are the atom-based Thole parameters that dictate the extent of the scaling between specific atom types. The use of atom-based Thole parameters has been shown to yield improvements in the treatment of the orientation of molecular polarizabilities.<sup>41</sup> In addition, the use of Thole screening has been extended to nonbond atom pairs.<sup>49</sup> This extension was motivated by the need to fine-tune interactions involving divalent ions, but the approach is general and may be applied to any atom pair.

When performing minimization and MD studies using a polarizable model, it is necessary, in principle, to perform a self-consistent field (SCF) calculation based on the Born–Oppenheimer approximation. This assumes that the electronic degrees of freedom relax for each position of the atomic nuclei. With the Drude model, this implies that the Drude particles relax in the electric field for each (fixed) nuclear configuration of the system. This requirement may be satisfied by performing an energy minimization of the Drude particles while the atomic nuclei positions are kept at a fixed position. However, this approach is computationally demanding, making it inappropriate for MD simulations. To overcome this, the electronic degrees of freedom are treated as classical dynamic variables in the MD simulation assuming an extended Lagrangian approach.<sup>71–73</sup> With the Drude model, this involves ascribing a mass to the Drude particles, which is taken from the parent atom and is typically 0.4 AMU. The extended Lagrangian procedure then allows the Drude particle to propagate classically during the simulation. A dual-thermostat Nose–Hoover algorithm based on the Velocity Verlet integrator was developed<sup>74</sup> in which the Drude particles are cooled to a very low temperature (near 0 K) during the MD simulation, thereby imposing the electronic degrees of freedom to approach the adiabatic SCF limit. Alternate dual thermostating approaches may be applied, notably based on Langevin dynamics, as has been implemented in the program NAMD.<sup>75</sup>

An issue when performing MD simulations using a polarizable force field is the potential for polarization catastrophe. This may occur when, for example, a positively charged ion approaches too close to an atom, leading the negatively charged Drude particle to become overpolarized, yielding unphysical infinite energies and instabilities. To avoid this issue, two approaches have been implemented. The first involves a “hyperpolarization” term where higher order anharmonic terms are added to the bond between the atomic core and the Drude particle.<sup>49</sup> Equation 4 shows the form of the hyperpolarization term

$$E_{hyp} = \sum K_{hyp}(R - R_o)^n \quad (4)$$

where  $R$  is the distance between the nucleus and the Drude particle,  $n$  is the order of the term, typically 4 or 6,  $K_{hyp}$  is the force constant, and  $R_o$  defines the distance at which the term starts to impact the Drude particle, typically 0.2 Å, such that the normal trajectory of the Drude is not impacted by the higher order term. More recently, a Drude reflective “hard wall” term has been added to more rigorously avoid a polarization catastrophe.<sup>53</sup> This term involves reversing the relative velocities along the Drude particle–parent atom nucleus bond and scaling them accordingly to the temperature of the Drude bond whenever the Drude–parent atom bond length exceeds a specified distance, again typically 0.2 Å. In addition, the relative displacement with respect to the specified distance

Table 1. Peptides and Proteins Subjected to MD Simulations<sup>a</sup>

protein	PDB ID/reference	box size (Å)	ions	simulation details
(Ala) <sub>5</sub>	--	31.8	--	Hamiltonian replica exchange
acetyl-(Ala) <sub>7</sub> -amide	--	--	--	gas-phase temp. replica exchange
(A) GB1 hairpin (residues 41–56 of protein G)	GEWTYDDATKTFVTE	50.0	--	150 ns MD with Drude FF
(B) dimeric coiled coil, 2 × 21 aa	1u0i <sup>132</sup>	58.7	0.15 M KCl	200 ns MD with C36 FF <sup>5</sup> 100 ns with Drude FF
(C) crambin, 46 aa	1ejg <sup>133</sup>	52.0	--	100 ns with C36 FF 2 × 100 ns with Drude FF
(D) protein GB3 domain, 56 aa	1p7e <sup>134</sup>	56.7	K <sup>+</sup>	100 ns with C36 FF 100 ns with Drude FF
(E) cold-shock protein A, 69 aa	1mjc <sup>135</sup>	52.4	K <sup>+</sup>	100 ns with C36 FF 100 ns with Drude FF
(F) ubiquitin, 76 aa	1ubq <sup>136</sup>	58.4	--	100 ns with C36 FF 100 ns with Drude FF
(G) circular permutant of ribosomal protein S6, 77 aa	3zzp <sup>137</sup>	61.0	Na <sup>+</sup>	100 ns with C36 FF 2 × 100 ns with Drude FF
(H) DNA methyltransferase associated protein (DMAP1)	4iej (to be published), 93 aa	59.0	--	100 ns with C36 FF 2 × 100 ns with Drude FF
(I) PDZ domain from tight junction regulatory protein, 94 aa	3vqf (to be published)	58.0	Na <sup>+</sup>	100 ns with C36 FF 2 × 100 ns with Drude FF
(J) lysozyme, 129 aa	135l <sup>138</sup>	69.8	Cl <sup>-</sup>	200 ns with C36 FF 93 and 100 ns with Drude FF
(K) fatty acid binding protein, 132 aa	1ifc <sup>139</sup>	63.4	--	100 ns with C36 FF 100 ns with Drude FF
(L) dethiobiotin synthase, 224 aa	1byi <sup>140</sup>	72.4	Na <sup>+</sup>	90 ns with C36 FF 2 × 90 ns with Drude FF

<sup>a</sup>-- indicates no ions or the periodic conditions were not used. Number of counterions added were enough to neutralize the system unless the concentration is presented.

is reversed and scaled according to the new velocities on the Drude particle to ensure that the location of the Drude is within the specified distance. In practical experience, the reflective hardwall term represents a more robust method to avoid polarization catastrophe as compared to the hyperpolarization terms and is now the recommended approach to be used on all MD simulations using the Drude polarizable force field.

Concerning the computational costs of the Drude implementation, as the model simply involves additional charged particles, the overhead is associated directly with the number of Drude particles. Thus, limiting the use of Drude particles to only non-hydrogen atoms limits the number of additional particles such that the increased computational cost is a factor of 1.6 to 2.0.<sup>75</sup> However, for many molecular systems, it is possible to use an integration time step of 2 fs with an additive force field, while the high frequency motions of the Drude particles limit the integration time step to 1 fs for the majority of systems, though a smaller time step may be required for highly ionic systems. In the present study, an integration time step of 1 fs was used, unless noted, such that the additional computational cost compared to the additive protein MD simulations is approximately a factor of 4.

## METHODS

**Quantum Mechanical Calculations.** Quantum mechanical (QM) calculations used the programs Gaussian 09<sup>76</sup> and Q-CHEM.<sup>77,78</sup> Energy minimizations were typically performed to default tolerances with the MP2/6-31G\* basis set with single point calculations at the RIMP2/cc-pVTZ level unless noted.<sup>79,80</sup> Interaction energies were calculated on rigid geometries at the RIMP2/cc-VQZ level without correction

for basis superposition error. The alanine, glycine, and proline dipeptide 2D  $\phi$ ,  $\psi$  energy surfaces were performed in 15° increments with optimization at the MP2/6-311G(d,p) level, followed by single point calculations at the RIMP2/cc-pVTZ and RIMP2/cc-pVQZ levels followed by extrapolation to the complete basis set (CBS) limit (MP2/6-311G(d,p)//RIMP2/CBS).<sup>81</sup>

**Empirical Force Field Calculations.** The Drude polarizable model was initially implemented in CHARMM allowing for a range of calculations<sup>82</sup> to be performed as required for force field optimization.<sup>83</sup> Subsequently, the model was implemented in NAMD and shown to be highly parallelizable in MD simulations.<sup>75</sup> The SWM4-NDP water model<sup>38</sup> was used in all calculations. Empirical force field calculations on the model compounds in the gas phase used no truncation of nonbond interactions, including dihedral potential energy scans. Minimizations were performed initially on only the Drude particles with the atoms constrained using the Steepest Descent (SD) optimizer. This was followed by minimization of all particles using the Adopted-Basis Newton–Raphson (ABNR) method to a gradient of 10<sup>-4</sup> kcal/mol/Å.

MD simulations were performed using the Velocity Verlet Integrator<sup>74</sup> in CHARMM and the Langevin Dynamics method implemented in NAMD.<sup>75</sup> Electrostatic interactions were treated using the particle mesh Ewald (PME) approach with a real space cutoff of 10 or 12 Å, a grid spacing of approximately 1 Å, and a fourth or sixth order spline. Lennard-Jones interactions were truncated at 10 or 12 Å with the smoothing performed over the final 2 Å using the switch algorithm<sup>84</sup> with the long-range LJ contribution accounted for using an isotropic correction.<sup>85</sup> All covalent bonds involving hydrogen as well as the intramolecular geometries of water were constrained using

the SHAKE<sup>86</sup> (CHARMM) or SETTLE<sup>87</sup> (NAMD) algorithms.

Simulation systems were initially prepared and equilibrated in the context of the CHARMM36 additive FF.<sup>5</sup> The protein structures were obtained from the protein database (PDB)<sup>88</sup> followed by solvation in a pre-equilibrated cubic TIP3P<sup>89</sup> water box of suitable size (Table 1) using the CHARMM-GUI.<sup>90</sup> Each system was then minimized for 200 SD steps with the protein constrained followed by minimization for another 200 SD steps without constraints on the protein. The systems were then heated to 300 K and subjected to a 100 ps NVT simulation followed by a 100 ps NPT simulation. The final coordinates were then regenerated in the context of the Drude polarizable force field. The Drude oscillators were minimized for 200 ABNR steps with all other atoms constrained; starting structures for Drude simulations were generated. Systems were then subjected to an additional 1 ns equilibration NPT simulation with a shorter time steps (1 fs for the additive and 0.5 fs for the Drude simulations) followed by the production simulations using NAMD, as detailed in Table 1. Coordinates were typically saved every 10 ps for analysis.

Hamiltonian replica exchange MD simulations<sup>91</sup> (H-REMD) were performed on  $(\text{Ala})_5$  in the NPT ensemble. The peptide was unblocked and had a protonated C terminus, corresponding to the experimental pH of  $\sim 2$ .<sup>92</sup> The H-REMD system consisted of eight replicas, where the potential function was perturbed using CMAP applied to the  $\phi$ ,  $\psi$  dihedrals. The perturbation involved inverting the CMAP surface with the perturbation involving scaling between the two CMAPs by factors of 0.1, where the ground state replica (replica0) uses the unperturbed CMAP and replica7 involves scaling of 0.7. The initial conformations of the peptide consisted of random coils in a periodic system including 988 water molecules in a cube of  $\sim 31.5$  Å/side. An in-house code interfacing NAMD and CHARMM was used to determine whether suitable exchanges had occurred. Each replica was simulated with NAMD for 2 ps at a constant pressure of 1 atm and a temperature of 300 K. Coordinates were saved each 1 ps. Simulations were performed with a 1 fs integration time step, with the exception of the N-methylacetamide (NMA) based model (Drude-1, see below), where a 0.5 fs time step was used. The Drude hardwall potential was not used during the simulations. After all replica simulations had terminated, CHARMM was used to evaluate the energy of each replica from the saved trajectories, and coordinates were exchanged if the energy of  $R_{i+1}$  was lower than the energy of  $R_i$ , where  $i$  indicates the replica number, or the exchange was accepted according to the Metropolis criterium.<sup>93</sup> The C5 region of the Ramachandran map is defined by the intervals  $-180^\circ < \phi < -90^\circ$  and  $50^\circ < \psi < 180^\circ$ ,  $160^\circ < \phi < 180^\circ$  and  $110^\circ < \psi < 180^\circ$ , and  $-180^\circ < \phi < -90^\circ$  and  $-180^\circ < \psi < -120^\circ$ . A residue is considered to belong to the PPII region if the  $\phi$  and  $\psi$  torsions fall in the following intervals:  $-90^\circ < \phi < -20^\circ$  and  $50^\circ < \psi < 180^\circ$  and  $-90^\circ < \phi < -20^\circ$  and  $-180^\circ < \psi < -120^\circ$ , and a residue is defined as right-handed alpha helical,  $\alpha R$  or  $\alpha +$ , if  $-160^\circ < \phi < -20^\circ$  and  $120^\circ < \psi < 50^\circ$  as well as a more rigorous definition, alpha\_h, where  $-100^\circ < \phi < -30^\circ$  and  $-67^\circ < \psi < -7^\circ$  with three consecutive residues falling in this region, as previously performed.<sup>62</sup>

Temperature REMD simulations<sup>94,95</sup> (T-REMD) were performed on the Ac-(Ala)<sub>7</sub>-amide polypeptide in the absence of explicit solvent using CHARMM with no cutoff of the nonbond interactions and a dielectric constant of 80. A total of 12 replicas were used with temperatures ranging from 300 to

600 K. Each window was simulated for 100 ns using Langevin dynamics with a friction coefficient of  $80 \text{ ps}^{-1}$ . Results are reported for the 300 K replica.

**Polypeptide Backbone Parameter Fitting.** The initial set of backbone parameters, referred to as Drude-1, were based on NMA.<sup>41,96</sup> Electrostatic parameters based on the alanine dipeptide, yielding the Drude-2 model, were determined by averaging the terms over five independent sets of parameters obtained from electrostatic potential (ESP) fitting corresponding to the  $\alpha R$ ,  $\alpha L$ , C5, PPII, and C7eq conformations. For each conformation, the electrostatic parameter optimization, which included the partial atomic charges, atomic polarizabilities, and atom-based Thole factors, was performed using the FITCH-ARGE module of CHARMM by fitting to the QM ESP maps as previously described.<sup>43,51</sup> The initial atomic charges were taken from the additive CHARMM model<sup>3</sup> and the initial atomic polarizabilities from the work of Miller.<sup>97</sup> The ESP grids were placed on concentric nonintersecting Connolly surfaces around the target molecule. Fitting targeted the “unperturbed” ESP for the molecule alone and the “perturbed” ESPs obtained by placing a point charge of  $+0.5e$  at different location around the molecule to probe the polarization response. The QM ESP maps were calculated using the B3LYP functional with the aug-cc-pVDZ basis set.

Additional optimization of backbone electrostatic parameters used a Monte Carlo Simulated Annealing (MCSA) protocol, yielding the final model, Drude-3.<sup>98</sup> The target data utilized an array of QM observables determined for the alanine dipeptide and large alanine polypeptides. Target data included the polarizability of the alanine dipeptide, relative energies of  $(\text{Ala})_5$ , and energetic and structural data for the interaction of the alanine dipeptide with individual water molecules along specific directions (Figure S1 of the Supporting Information). Several conformations of the alanine models were used in the development of the Drude-3 model: (i)  $\alpha R$ , C5, and PPII for the relative energies of  $(\text{Ala})_5$ ; (ii) C5 and PPII for the interactions of the alanine dipeptide with water; and (iii)  $\alpha R$ , C5, PPII, and C7eq conformations of the alanine dipeptide for molecular polarizabilities and dipole moments. In addition to the electrostatic parameters, during the MCSA, internal parameters were allowed to vary within a limited range to keep the alanine dipeptide optimized geometries close to the targeted values. MCSA started with a temperature of 150 K with individual parameters randomly adjusted followed by accepting or rejecting the new parameter set based on the Metropolis criterium.<sup>93</sup> The temperature was gradually reduced to near 0 K, yielding a selected parameter set for testing in  $(\text{Ala})_5$  solution simulations. The error function is the sum of all differences between MM and QM data for all properties mentioned above with various weighting factors. During MCSA fitting, a new CMAP that reproduces the QM alanine dipeptide  $\phi$ ,  $\psi$  MP2/6-311G(d,p)//RIMP2/CBS surface was generated at each iteration. In addition, adjustments of the CMAP away from the QM-based surface to improve agreement with conformational sampling of the peptide backbone in peptides and proteins were included in the final Drude-3 model.

**$\chi_1$  and  $\chi_2$  Dihedral Parameter Fitting.** Side chain  $\chi_1$  and  $\chi_2$  dihedral parameters for amino acids, excluding Ala, Gly, and Pro, were initially fitted to QM potential energy surfaces for the  $\chi_1$  and  $\chi_2$  torsions in amino acids capped with N-acetyl and N'-methylamide moieties, analogous to the alanine dipeptide. Previously published 2D scans of  $\chi_1$  and  $\chi_2$  dihedral angles were performed in  $15^\circ$  increments with dipeptides constrained at

three different backbone conformations,  $\alpha R$  ( $-60.0, -45.0$ ),  $\beta$  ( $-120.0, 120.0$ ), and  $\alpha L$  ( $63.5, 34.8$ ), at the RIMP2/cc-PVTZ//MP2/6-31G\* (6-31+G\* for the charged Asp and Glu dipeptides) level, and the resulting three 2D energy surfaces were offset relative to the global minimum for each amino acid type.<sup>99</sup> This set of QM surfaces was also used as the target data for the initial optimization of the side-chain  $\chi_1$  and  $\chi_2$  parameters for the additive CHARMM36 force field.<sup>5</sup>

CHARMM was used to generate analogous 2D MM energy surfaces, and a MCSA automated fitting program<sup>100</sup> was used to minimize the weighted root-mean-square (RMS) energy difference between the MM and QM surfaces for the optimization of side chain dihedral parameters. During the fitting, a higher weighting factor of 5 was assigned for energies from  $\alpha R$  and  $\beta$  backbone conformations versus 1 for those from  $\alpha L$ . An energy cutoff of 12 kcal/mol was applied, and all the energies higher than the cutoff were discarded to avoid fitting these high-energy regions at the expense of distorting the low energy regions. For charged side chains, higher energy cutoffs (20 kcal/mol for Arg and Lys and 25 kcal/mol for Asp, Glu, and His) were used. Amino acids sharing the same  $\chi_1$  and  $\chi_2$  parameters (Lys/Arg/Met, Tyr/Phe, and Thr/Ile/Val) were fitted together. During the MCSA optimization of dihedral parameters, the multiplicities ( $n$ ) were limited to the combination of 1, 2, and 3, while the force constants ( $K$ ) were upper-bounded to 4.0 kcal/mol and phases ( $\delta$ ) were restricted to either 0 or 180°. Simulated annealing starting from an initial temperature of 1000 K was carried out for 200 000 steps with exponential cooling.

Selected  $\chi_1$  and  $\chi_2$  dihedral parameters optimized to target the dipeptide gas phase QM surfaces were subjected to manual adjustment based on condensed phase simulations. For each amino acid X, the 9-mer peptide (Ala)<sub>4</sub>—X—(Ala)<sub>4</sub> was solvated in a 32 Å cubic water box with the backbone restrained to the C7eq conformation ( $-82.8^\circ, 77.9^\circ$ ) conformation. A special Hamiltonian replica exchange with the solute scaling method (REST2)<sup>101</sup> was used to enhance sampling efficiency. This is a revision of the solute tempering replica exchange method proposed by the same group<sup>102</sup> and is based on the decomposition of the total interaction energy into peptide intramolecular energy ( $E_{pp}$ ), peptide–water interaction energy ( $E_{pw}$ ), and self-interaction energy within the water molecules ( $E_{ww}$ ), and scaling these three potential energy component in such a way that the water–water energy vanishes from the REMD acceptance criterion. To be specific, all replicas are run at the same temperature  $T_0$  with the potential energy of the replica  $m$  given by

$$E_m = \frac{T_0}{T_m} E_{pp} + \sqrt{\frac{T_0}{T_m}} E_{pw} + E_{ww} \quad (5)$$

Only the 0th (ie. ground state) replica corresponds to the desired equilibrium distribution at  $T_0$ . It should be noted that  $T_m$  in eq 5 is not a simulation temperature but should be understood as a general scaling factor applied to the Hamiltonian. An in-house implementation of REST2 in CHARMM was used to run 10 ns solute scaling simulations with four replicas ( $T_0 = 300$  K,  $T_1 = 329$  K,  $T_2 = 363$  K, and  $T_3 = 400$  K) and an exchange attempting frequency of 0.1 ps<sup>-1</sup>. Coordinates were saved every 1 ps, from which  $\chi$  dihedral angles were extracted.

To assess the agreement between  $\chi_1$  and  $\chi_2$  distributions from MD simulations and from a crystallographic survey,<sup>99</sup>

probability histograms were generated using a bin size of 15° and the 1D overlap coefficient (OC) between two probability distributions was calculated as

$$OC = \frac{\sum p_m \cdot p_n}{\sqrt{\sum (p_m)^2 \cdot \sum (p_n)^2}} \quad (6)$$

and extended to two dimensions<sup>103</sup> in the present study.

## RESULTS AND DISCUSSION

Most additive force fields for biological polymers such as DNA and proteins were developed by first obtaining optimal parameters for a collection of small molecules representative of the chemical functionalities in the biomolecules which serve as building blocks.<sup>1,3,104</sup> These parameters were then “combined” together to yield the biopolymer force field. In this transfer, the connectivities between the various building blocks were optimized to yield conformational properties in satisfactory agreement with experimental data. Working from a set of small model compounds was critically important for the optimization of the nonbond parameters targeting experimental data, such as heats of vaporization or sublimation, densities or lattice volumes, and free energies of solvation, for individual types of functional groups. This approach worked well, and the earlier additive force fields, such as CHARMM22,<sup>3</sup> OPLS,<sup>7</sup> AMBER,<sup>1</sup> and GROMOS,<sup>105</sup> were subsequently successfully applied in a large number of application studies on biological macromolecules. The successful transfer of the small compound parameters appears to be due in part to the additive nature of the force field, where combining the “parts” into the complete biopolymers yielded satisfactory behavior in macromolecules. However, it should be noted that additional refinement of the additive force fields has been ongoing for over 20 years; our own experience with an all-atom protein force field started around 1988 and saw completion of a first generation model in May 1992 (CHARMM22), though not published until 1998,<sup>3</sup> followed by the CMAP extension in 2002,<sup>4,106</sup> and the most recent iteration to yield CHARMM36 completed in 2012.<sup>5</sup>

**Polypeptide Backbone Parameter Optimization.** Development of the Drude protein force field was initially performed following an approach similar to that for the additive model. Thus, the initial model of the polypeptide backbone in the form of polyalanine, the Drude-1 model, was based on a combination of parameters from NMA and ethane, with the links between those moieties adjusted to reproduce QM and protein crystal survey data. The NMA parameters are a recently revised version that yielded good agreement with a variety of condensed phase properties as well as giving hydrogen bond distances consistent with those occurring in  $\alpha$  helices.<sup>96</sup> Importantly, in creating the initial NMA-based polypeptide model, the  $(\phi, \psi)$  Ramachandran map was tuned to reproduce a high-level QM (RIMP2/cc-pVQZ//RIMP2/CBS) surface using the CMAP utility in CHARMM, where the CBS (complete basis set) extrapolation<sup>81</sup> was obtained based on RIMP2/cc-VTZ and RIMP2/cc-VQZ single point energies. Analogous QM surfaces were used for CMAP terms for Gly and Pro and used without further adjustments. Initial validation of the Drude-1 model was performed using MD simulations of (Ala)<sub>5</sub> in solution,<sup>92</sup> a system that serves as an important benchmark for protein force fields.<sup>57</sup> However, as shown in Table 2, the agreement with the NMR J-coupling data was poor. Analysis of the  $\phi, \psi$  distribution shows that this was due to the peptide predominantly populating extended conforma-

**Table 2. Comparison of Experimental and Calculated NMR  $^3J$  Couplings for and the Percent Secondary Structures for  $(\text{Ala})_5$  in Aqueous Solution<sup>a</sup>**

		$\chi^2$	$\alpha+$	C5	PPII
Drude-1	average	6.3	0.1	99.6	6.3
	std error	0.1	0.1	0.1	0.1
Drude-2	average	4.3	0.3	89.0	10.5
	std error	0.3	0.2	2.2	2.1
Drude-3	average	2.3	2.1	72.9	24.6
	std error	0.3	0.6	2.8	2.3
C36	ref 5	1.2	13.1	30.8	51.9
			1.3	1.3	1.1

<sup>a</sup> $\chi^2$  values calculated following the approach of Best and co-workers<sup>57</sup> with values averaged over residues 2 through 4. Statistical analysis from block error analysis using 2 ns blocks from 2–10 ns with Drude-1 and 2–12 ns with Drude-2 and -3.

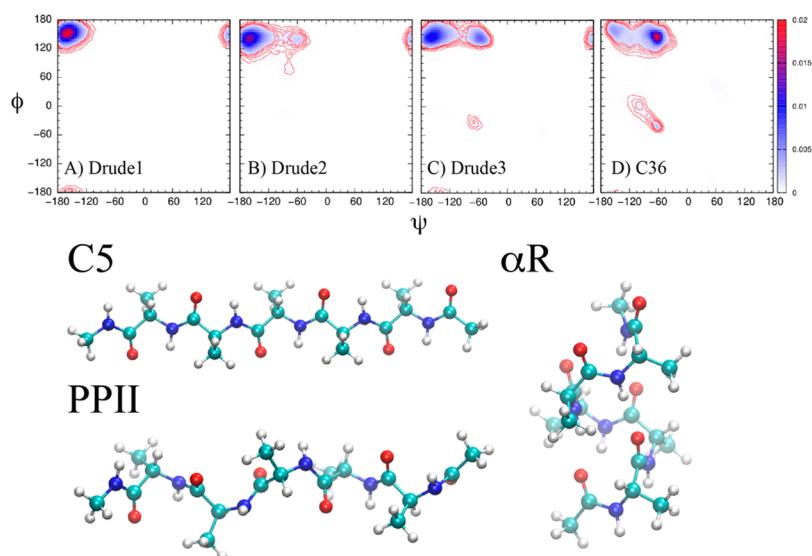
tions (Figure 2). Included in Figure 2 are three representative conformations of blocked polyalanine in the extended (ie. C5 or beta), PPII, and  $\alpha$ R states. Notable is the difference in the C5 and PPII states, where in the C5 conformation the adjacent peptide bonds are coplanar and aligned in opposite directions. This had two direct effects. First, it allowed for maximal hydrogen bonding with the aqueous environment, and in the polarizable model, the local dipoles associated with the individual peptide bonds interact with each other to enhance the local dipole moments associated with each peptide bond. Indeed, a comparison of the dipole moments of acetyl-(Ala)<sub>5</sub>-amide for the NMA based model with QM data indicated the overall dipole moment of the C5 conformation to be significantly overestimated (Table 3). It was therefore hypothesized that the overestimation, which would lead to even more favorable interactions with aqueous solvent, was due to the electrostatic parameter optimization procedure based on NMA alone not taking into account the electrostatic communication between the individual peptide bonds.

**Table 3. Gas Phase Dipole Moments (Debye) of Acetyl-(Ala)<sub>5</sub>-NHCH<sub>3</sub>**

	RIMP2/cc-pVTZ	B3LYP/6-31G(d)	C36	Drude-1	Drude-2	Drude-3
C7eq	13.4	11.3	8.3	2.8	16.9	11.2
C5	10.2	11.6	8.8	24.4	4.5	9.2
PPII	7.1	8.1	2.9	13.3	3.0	2.4
$\alpha$ R	24.5	22.0	21.9	13.5	22.4	21.1

On the basis of this analysis, it was concluded that a model compound allowing communication between adjacent peptide bonds was required, with the initial candidate being the alanine dipeptide. Five conformations of the alanine dipeptide were selected covering different relative orientations of the peptide bonds (C7eq, C5, PPII,  $\alpha$ R, and  $\alpha$ L). The five conformations were subjected to constrained QM optimizations and subjected to RESP fitting yielding five electrostatic models, which were then averaged. This produces electrostatic parameters that better reproduce the change in the ESP associated with electrostatic interactions between the peptide bonds in the different relative orientations. The resulting Drude-2 model yielded a smaller dipole moment for the C5 conformation for Acetyl-(Ala)<sub>5</sub>-amide (Table 3) and was then used in simulations of  $(\text{Ala})_5$ , with the results included in Table 2 and Figure 2. As compared to the Drude-1, the NMA-only based model, the Drude-2 model shows improved agreement with experimental results, though the agreement is still poor as compared to C36. Notably, the PPII region is populated (Figure 2b), though the C5 conformation still dominates, indicating that the inclusion of electrostatic interactions between the peptide bonds during parameter optimization did improve the situation. However, those improvements were clearly insufficient, indicating that additional target data were needed to obtain a more accurate electrostatic model for the polypeptide backbone.

Subsequent efforts leading to the final Drude-3 model accounted for the contribution of the electrostatic term to the



**Figure 2.**  $\phi, \psi$  probability distribution from the  $(\text{Ala})_5$  simulations using electrostatic parameters based on (A) N-methylacetamide (Drude-1), (B) ESP fitting to multiple conformations of the alanine dipeptide (Drude-2), (C) fitting to the target data in Table 2 (Drude-3) with adjustments to relative energies of different regions of the CMAP  $\phi, \psi$  surface, and (D) from the CHARMM36 additive force field. The bottom portion of the figures shows images<sup>130</sup> of the acetyl-(Ala)<sub>5</sub>-methylamide, the C5, PPII, and  $\alpha$ R conformations.

**Table 4.** Target Data for Monte Carlo Simulated Annealing Optimization of the Polypeptide Backbone Electrostatic Parameters along with the Values from the Final Drude-3 Model<sup>a</sup>

		relative energies of acetyl-(Ala) <sub>5</sub> -amide								
		RIMP2/cc-pVDZ		RIMP2/cc-pVTZ//RIMP2/cc-pVDZ		Drude-3				
alphaR - C5		-10.39		-6.59		-3.89				
alphaR - PPII		-16.63		-14.83		-10.17				
water interactions with alanine dipeptide <sup>b</sup>										
		C5				PPII				
		int. energy		distance		int. energy		distance		
		QM	Drude-3	QM	Drude-3	QM	Drude-3	QM	Drude-3	
OY-LP		-6.48	-6.35	1.89	1.89	-6.07	-5.78	1.90	1.91	
O-LP		-5.60	-5.13	1.93	1.95	-6.13	-5.54	1.91	1.94	
OY-straight		-5.50	-5.71	1.93	1.95	-5.80	-5.24	1.92	1.95	
O-straight		-4.71	-4.75	1.94	1.95	-5.26	-5.08	1.93	1.95	
HN		0.50	0.40	3.01	3.00	-5.15	-5.90	1.96	1.92	
HNT		-5.68	-6.18	2.03	1.99	-5.76	-6.40	1.97	1.91	
molecular dipole moment of the alanine dipeptide (Debye) <sup>c</sup>										
$\alpha R$		C5		PPII		C7eq				
		QM	Drude-3	QM	Drude-3	QM	Drude-3	QM	Drude-3	
$\mu$	6.23	6.67	4.66	2.63	4.17	3.59	2.32	3.77		
$\mu_x$	1.32	3.02	-4.40	-2.27	-0.93	0.82	2.29	3.62		
$\mu_y$	-1.62	-1.48	-1.01	-0.28	-3.67	-3.33	-0.27	-0.70		
$\mu_z$	5.87	5.76	1.18	1.31	-1.74	-1.09	0.30	0.78		
molecular polarizability of alanine dipeptide ( $\text{\AA}^3$ )										
$\alpha R$		C5		PPII		C7eq				
		QM	Drude-3	QM	Drude-3	QM	Drude-3	QM	Drude-3	
$\alpha_{xx}$	13.57	15.30	15.49	16.07	14.79	16.39	14.06	15.47		
$\alpha_{yy}$	12.72	14.36	12.06	12.78	12.25	13.88	13.92	14.69		
$\alpha_{zz}$	11.71	9.94	10.35	10.39	11.01	9.45	10.40	9.72		

<sup>a</sup>Energies in kcal/mol and distances in Å. <sup>b</sup>Water interactions correspond to the minimum interaction energy distances with rigid monomer geometries with the QM data obtained using MP2/aug-cc-pVDZ optimizations followed by single point energies at the MP2/cc-pVQZ model chemistry. <sup>c</sup>QM dipole moments and polarizabilities obtained at the B3LYP/aug-cc-pVDZ level with the polarizabilities scaled by 0.85 prior to fitting.

intramolecular electrostatic properties and conformational energies as well as to the intermolecular interactions with the environment (Table 4). It was hypothesized that reproduction of relative gas phase energies of acetyl-(Ala)<sub>5</sub>-amide together with including interactions with a water molecule in the fitting procedure would help achieve the correct relative energies in aqueous solution by balancing the electrostatic contribution to both the intra- and intermolecular energies. To obtain target data for the intramolecular energies, we undertook QM calculation on acetyl-(Ala)<sub>5</sub>-amide for the C5 (-158.5, 161.6), PPII (-75, 150), and  $\alpha R$  helical (-60, -45) conformations. From these calculations, the relative energies of the C5- $\alpha R$  and PPII- $\alpha R$  conformations were used as target data. To weight the intrinsic quality of the electrostatic model itself, QM data on the dipole moments and the components of the polarizability tensor of the C5, C7eq, PPII, and  $\alpha R$  conformations of the alanine dipeptide were included as target data. The QM polarizabilities were scaled by 0.85, consistent with scaling used during previous development of the Drude small compound parameters.<sup>17,47</sup> Finally, consideration of intermolecular energies in the optimization was taken into account by including interactions of water with the alanine dipeptide in the C5 and PPII conformations (Figure S1 of the Supporting Information). On the basis of these target data, a series of MCSA runs were carried out, with varying weights on

the different target data. This generated a collection of electrostatic parameter sets for testing.

Testing of the electrostatic parameters initially involved MD simulations of (Ala)<sub>5</sub> in solution. From these simulations, several parameter sets with satisfactory agreement with the NMR J-coupling data were selected. Simulations were then undertaken using these models on acetyl-(Ala)<sub>5</sub>-amide in the gas phase using Langevin dynamics with the dielectric set to 80 in the context of T-REMD to gauge their ability to assume helical structures, using simulations of this system with the C36 force field as a benchmark. From this process, several sets of parameters were selected and further tested by carrying out MD simulations of crambin, lysozyme, and the GB1 hairpin peptide using preliminary side chain  $\chi_1$  and  $\chi_2$  parameters (see below). At this stage, the  $\phi$ ,  $\psi$  distributions were monitored along with RMS differences for the two proteins and the peptide and J-coupling data for (Ala)<sub>5</sub>. Based on these data, adjustments were made to the 2D CMAP energy term to improve the  $\phi$ ,  $\psi$  sampling in the context of (Ala)<sub>5</sub>, GB1, acetyl-(Ala)<sub>5</sub>-amide, crambin, and lysozyme. From this process, the final Drude-3 model was selected, with the remainder of the results presented below based on the Drude-3 model. This model will be subsequently referred to as the “Drude-2013” force field for public release. We note that a number of the tested models yielded improved agreement with the NMR data for (Ala)<sub>5</sub> (not shown) but significantly underestimated the  $\alpha R$

content and/or had unsatisfactory sampling of  $\phi, \psi$  in lysozyme and crambin. The recent AMOEBA-13 model yielded good agreement with the NMR data for (Ala)<sub>5</sub> with a  $\chi^2$  value of 1.0 for all residues, although the amount of helical sampling in acetyl-(AAQAA)<sub>3</sub>-amide is underestimated based on 30 ns TREMD simulations.<sup>36</sup> Thus, the Drude-3 model represents a compromise between these different target data obtained in the context of limited simulation times, a compromise that leads to the Drude model being in poorer agreement with the experimental data as compared to C36 with respect to the reproduction of the (Ala)<sub>5</sub> NMR data. This poorer agreement is associated with enhanced population of the C5 conformation over PPII, which is reversed in the additive C36 model. Furthermore, we note that the helical propensity of the model is not well determined. Based on acetyl-(Ala)<sub>7</sub>-amide (Table 5),

**Table 5. Secondary Structural Content of Acetyl-(Ala)<sub>7</sub>-amide from Langevin T-REMD Simulations with a Dielectric Constant of 80<sup>a</sup>**

force field	% $\alpha$ R	%C5	%PPII
Drude-1	3.8 ± 0.1	84.9 ± 0.1	0.5 ± 0
Drude-2	7.8 ± 0.2	9.9 ± 0.1	52.2 ± 0.4
Drude-3	62.0 ± 1.0	20.8 ± 0.6	15.5 ± 0.3
CHARMM36	28.1 ± 0.6	25.5 ± 0.4	9.2 ± 0.3

<sup>a</sup>% $\alpha$ R represents the percentage of residues occupying the helical region. Averages and standard errors based on 5 blocks of 20 ns.

the Drude-3 model has significantly more helical content than the Drude-1 and Drude-2 models as well as C36. However, given the polarizable nature of the Drude force field, it is hard to gauge if the model will have significantly more helical character than C36 in aqueous solution. For example, based on (Ala)<sub>5</sub>, C36 yields more helical configurations than the Drude-3 model (Table 2). Attempts to determine the helical propensity of the model using acetyl-(AAQAA)<sub>3</sub>-amide, as was done with C36, were unsuccessful due to inadequate convergence. Reported results for C36 used 140 ns of sampling for each replica in T-REMD simulations, with the helical content determined over the final 40 ns,<sup>62</sup> computational time scales that are currently difficult to attain with the polarizable Drude model. Accordingly, we anticipate that future optimization of the Drude polarizable model based on a wider range of target data, including improved sampling in condensed phase simulations, will be necessary.

#### Side Chain $\chi_1, \chi_2$ Dihedral Parameter Optimization.

Side chain identity is known to impact the conformational distribution of the polypeptide backbone, as indicated by the extensive experimental studies of Baldwin and co-workers.<sup>107–109</sup> Furthermore, while the conformational properties of the individual side chains themselves will impact backbone conformational properties, the conformation of the backbone is known to influence side chain conformations as well.<sup>110–112</sup> Thus, this linkage between the conformational properties of the backbone and the side chain conformational properties indicates that an iterative optimization approach is required. To overcome this, we investigated the use of the peptide (Ala)<sub>4</sub>X(Ala)<sub>4</sub> as a model system for  $\chi_1, \chi_2$  parameter optimization,<sup>113</sup> where X is the amino acid of interest, and the backbone conformation was constrained to fully extended, C7eq, PPII, and  $\alpha$ R conformations.<sup>114</sup> These studies indicated that (Ala)<sub>4</sub>X(Ala)<sub>4</sub> with either the C7eq or PPII backbone conformation yields aqueous phase conformational properties

that mimic those occurring in full proteins. Based on this analysis,  $\chi_1, \chi_2$  parameter optimization was performed by initially targeting QM data for the respective side chain dipeptides, with the backbone in the  $\beta, \alpha$ R, and  $\alpha$ L conformations. These parameters were then used in H-REMD simulations of (Ala)<sub>4</sub>X(Ala)<sub>4</sub> in solution, with  $\chi_1, \chi_2$  sampling compared to the PDB survey data.

Presented in Table 6 are the overlap coefficients for  $\chi_1$  and  $\chi_2$  from the (Ala)<sub>4</sub>X(Ala)<sub>4</sub> in the C7eq conformation with

**Table 6. Impact of Condensed Phase Optimization on the Overlap of the (Ala)<sub>4</sub>X(Ala)<sub>4</sub> and PDB  $\chi_1/\chi_2$  Probability Distributions**

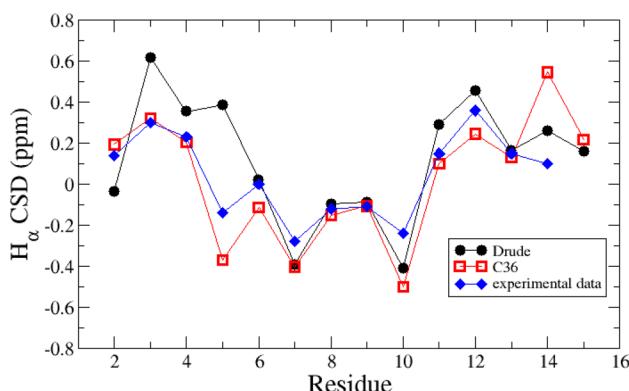
residue, X	$\chi_1$ OC		$\chi_2$ OC	
	QM only	final	QM only	final
Arg	0.56	0.82	0.70	0.85
Asn	0.75	0.82	0.51	0.61
Asp	0.54	0.65	0.58	0.79
Cys	0.87			
Gln	0.90	0.92	0.36	0.60
Glu	0.85	0.68	0.26	0.81
Hsd	0.80	0.94	0.77	0.80
Ile	0.07	0.78	0.15	0.89
Leu	0.88		0.92	
Lys	0.70	0.81	0.20	0.93
Met	0.80	0.92	0.75	0.88
Phe	0.69	0.94	0.87	0.89
Ser	0.47	0.74		
Thr	0.31	0.74		
Trp	0.73	0.86	0.59	0.74
Tyr	0.94	0.72	0.73	0.74
Val	0.87			

those from a survey of the PDB.<sup>114</sup> The extent of overlap for many of the amino acids based on optimization only targeting the QM data is quite good. For example, the values of 0.87, 0.88, and 0.87 were obtained for  $\chi_1$  for Cys, Leu, and Val, respectively, while the OC was 0.92 for  $\chi_2$  with Leu. Based on the quality of the fit for these residues, additional optimization was not performed. Additional optimization for the remaining residues involved comparison of the computed and target  $\chi_1$  and  $\chi_2$  populations of the gauche+, gauche-, and trans rotamers and manually adjusting the corresponding dihedral parameters to improve the level of agreement. To be more specific, the local minimum energies at the trans and gauche- conformations were shifted empirically by adjusting the 1- and 2-fold dihedral force constants. Since solute scaling replica exchange simulations require a minimal number of replicas, multiple parameter sets can be tested simultaneously. This iterative approach, which typically required ~9 days using eight cores per replica for the H-REMD calculations per iteration, led to significant agreement with the PDB target data for a number of amino acids, notable examples being Ile, Lys, and Thr. Overall, the final OC values are typically 0.7 or higher, though lower values are present including Asn  $\chi_2$ , Asp  $\chi_1$ , Gln  $\chi_2$ , and Glu  $\chi_1$ . The final parameters were used for the reported polypeptide and protein simulations.

**Peptide Simulations.** In addition to (Ala)<sub>5</sub>, two other peptides were examined through MD simulations, the GB1 (41–56) hairpin<sup>115,116</sup> and a dimeric coiled coil (1UOI).<sup>117</sup> Both peptides were used in a previous force field optimization study<sup>5</sup> and were included as part of the training set during final

optimization of the backbone parameters in the present study. Accordingly, these simulations do not provide a true validation, though the ability of the Drude model to reproduce the corresponding experimental properties may be considered an indicator of the quality of the model for the treatment of small, partially disorder peptides.

Calculations of the GB1 hairpin simply involved explicit solvent MD simulations using both the additive and Drude models. RMS difference analysis shows the Drude model to stay closer to the conformation of the hairpin observed in the crystal structure of the full GB1 protein (Figure S2a). The additive simulation drifts away from the crystal conformation, with the RMS difference fluctuating primarily between 2.5 and 3 Å. In contrast, the Drude simulation maintains conformations closer to the crystal structure on the time scale of the simulation, yielding conformations approximately 0.8 Å and 1.5–2.0 Å RMS away from the crystal conformation. The high occurrence of both highly native-like and non-native conformations is consistent with the midpoint of the folding transition being 297 K,<sup>116</sup> while a more recent study indicates the peptide to be 30% folded at 298 K.<sup>118</sup> Additional analysis involved calculation of the chemical shifts averaged over the simulations using Sparta+.<sup>119</sup> Presented in Figure 3 are the



**Figure 3.** Comparison of experimental and calculated NMR  $\text{H}\alpha$  chemical shifts for the GB1 hairpin for the Drude (black circles) and C36 (red squares) force fields. Calculated values were from Sparta+, and the experimental data correspond to 280 K, while the simulations were performed at 298 K.

shifts from the two simulations, performed at 298 K, along with the experimental data from 280 K.<sup>118</sup> The agreement of both the additive and Drude model is similar, with the additive model in better agreement for residues 2–5, while the Drude model shows better agreement with several residues later in the sequence. These results indicate that the additive model may better represent conformations of Trp3 and Tyr5, which are important for the hydrophobic interactions that contribute to stabilization of the structure, though additional sampling is required to more robustly compare the two models.

A simulation of a dimeric coiled-coil in solution with 150 mM KCl was initiated from model 1 of the NMR structures in 1UOI,<sup>117</sup> using a previously equilibrated system.<sup>5</sup> Analysis involved RMS differences,  $\phi,\psi$  distribution, and the interhelical angle. RMS analysis showed the overall structure of the complex to deviate more from model 1 of the NMR structure for the simulation based on the Drude-3 model versus C36 (Table 7 and Figure S2b of the Supporting Information). This was associated with the individual helices moving relative to

each other (Figure 4a), which occurs in both FFs, while the conformations of the individual helices were well preserved, indicating the ability of the model to adequately treat these essential types of secondary structure. This is supported by the  $\phi,\psi$  probability distribution from the simulations (Figure S3b, Supporting Information). Additional analysis involved the angles between the two helices. Both the additive and polarizable models sample a range of interhelical angles, spanning the range of values reported for the 20 models from the original NMR study. The Drude model populates a slightly wider range of angles during the 100 ns simulation, which is an indication of the additional flexibility in the model as further evidenced in the simulations of larger proteins, as presented below.

**Simulations of Full Proteins.** To further validate the Drude force field, simulations were performed on 10 additional proteins, as listed in Table 1 and shown in Figure S4, Supporting Information. The monomers of these proteins were subjected to ~100 ns simulations using both the polarizable and additive C36 models, with two simulations performed with the polarizable FF in select cases. The proteins were selected on the basis of coverage of secondary structure, experimental resolution, and the availability of results from previous studies. A majority of the proteins were small, being less than 100 amino acids (aa), as they would more robustly test the ability of the force field to maintain their folded structures as compared to large, globular proteins, as well as for computational expediency.

RMSD analysis was performed on all the proteins based on the  $\text{C}\alpha$  atoms following alignment of residues in the secondary structural elements or of all residues, with the results summarized in Table 7, and the RMSD time series for the individual systems are shown in Figure S2 of the Supporting Information. Table 7 also includes the RMS fluctuations of the RMSD time series. Two general trends are evident. The RMS differences are consistently smaller with the additive force field as are the RMS fluctuations. This indicates that the Drude model has additional flexibility as compared to the additive model. The only exception is the results for all residues in ubiquitin (1UBQ), where the average and fluctuations of the RMSD for all residues are larger in C36. This difference is primarily due to the five C-terminal residues, which have large B factors in the 1UBQ crystal structure.

Three of the proteins simulated in the present work were also studied in the recent AMOEBA protein FF paper, including crambin (1EJG), ubiquitin (1UBQ), and lysozyme (135L in the present work vs 6LYT in the AMOEBA study).<sup>36</sup> At the end of the 30 ns simulations in the AMOEBA study, the backbone RMSDs were approximately 1, 2, and 2 Å for the three proteins, respectively. These values compare to 1.1/1.1, 1.9 and 1.9/2.2 Å for those proteins at 30 ns of the present simulations, where two values are from the individual simulations of crambin and lysozyme. Thus, the two polarizable models yield similar RMS differences for 30 ns simulations.

While the Drude model exhibits larger flexibility as compared to the additive C36 model, the secondary structures are stable and maintained (Figure S3, Supporting Information). This suggests that the hydrogen bonding associated with secondary structure is being satisfactorily modeled. To verify this suggestion, the hydrogen bonding interactions involving the peptide bond N···O distance probability distributions in the helical and sheet secondary structure regions were analyzed. Shown in Figure 5 are the N···O probability distributions

**Table 7.** Average and RMS Fluctuations of the RMS Differences with Respect to the Crystal or NMR Structures of the Ca Atoms for the Residues in the Secondary Structures and for All Residues (Values in Å)<sup>a</sup>

protein		secondary				all residues			
		average		RMS fluc.		average		RMS fluc.	
		C36	Drude	C36	Drude	C36	Drude	C36	Drude
A	GB1					2.70	1.28	0.49	0.47
B	IU01	1.38	1.81	0.27	0.40	1.59	2.09	0.31	0.41
C	1EJG	0.81	1.08	0.09	0.10	0.99	1.19	0.14	0.13
				1.09	0.10		1.21		0.13
D	1P7E	0.56	1.31	0.10	0.19	0.65	1.72	0.12	0.31
E	1MJC	0.51	0.57	0.08	0.09	1.56	1.65	0.19	0.26
F	1UBQ	0.72	0.95	0.11	0.14	2.17	1.67	0.69	0.31
G	3ZZP	0.98	1.33	0.11	0.15	1.48	1.64	0.15	0.19
				1.46	0.33		1.80		0.33
H	4IEJ	0.69	1.33	0.07	0.40	1.07	1.63	0.11	0.39
				1.72	0.52		2.12		0.56
I	3VQF	1.01	1.08	0.11	0.14	1.17	1.91	0.15	0.27
				1.27	0.27		1.73		0.29
J	135L	1.34	1.60	0.19	0.44	1.53	1.97	0.19	0.42
				1.59	0.31		2.19		0.34
K	1IFC	0.87	1.78	0.10	0.27	1.05	1.92	0.12	0.25
L	1BYI	1.22	1.51	0.25	0.24	1.33	2.30	0.22	0.34
				1.30	0.18		2.33		0.49

<sup>a</sup>Averages over 100 ns, with the exception of 135L (93 and 100 ns) and 1BYI (100 ns). The second row for C, G, H, I, J, and L is for the second Drude simulation with those systems.

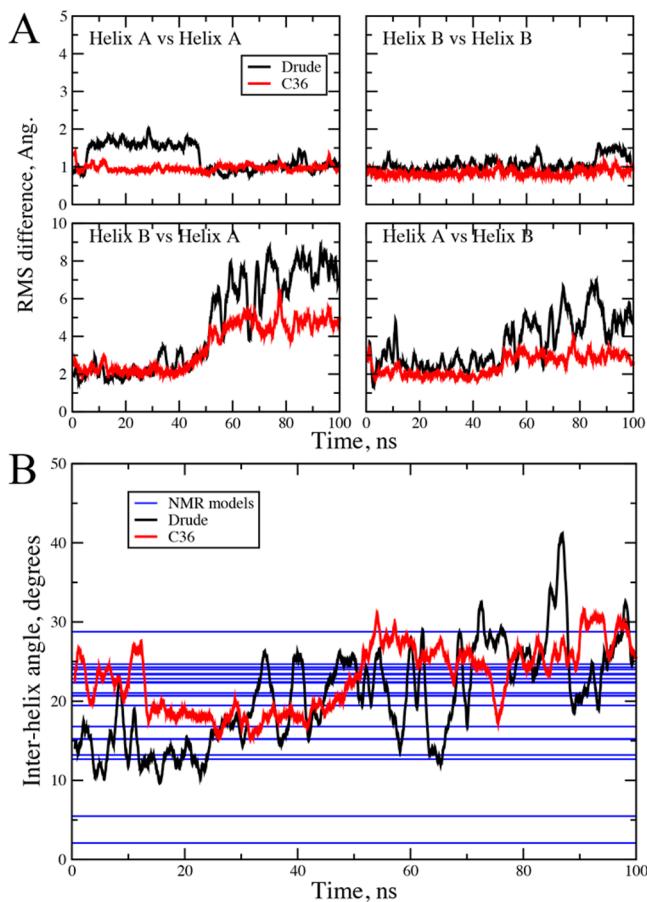
obtained from a subset of the simulated proteins along with survey data from crystal structures from the PDB with resolution higher than 1.5 Å.<sup>99</sup> With the helices, in the additive force field, the distribution goes to shorter distances than observed in the survey while the agreement of the Drude model for the leading edge of the curve with the Xray data is excellent, although the maximum in the Drude model is slightly shifted back from the survey by ~0.05 Å. This difference is larger with N···O distances in the sheets, with both the additive and polarizable models having maxima longer than that in the survey, with the difference being larger in the Drude model. Thus, the Drude model gives systematically longer N···O distances than the additive model, with agreement in helices somewhat better with the Drude model, while the opposite occurs with the sheet regions. It is important to recall that the shorter distances with the additive model are consistent with the optimization of the force field, where short hydrogen bonding distances were required to yield pure solvent properties for model compounds in agreement with experimental results (enthalpy and density).<sup>10</sup> Such shortened hydrogen bond distances are not required to as great an extent with the Drude model (Table 4); however, the present results indicate that a small, systematic decrease in the distances may be required in future generations of the Drude force field.

Further investigation into the structural properties of the polarizable model involved analysis of the  $\phi$ ,  $\psi$  distributions in the studied proteins. Presented in Figure 6 are the  $\phi$ ,  $\psi$  inverted Boltzmann weighted distributions over protein simulations C through L reported in Table 1, with the distributions for the individual systems shown in Figure S3 of the Supporting Information. Figure 6 also contains a distribution obtained from the survey of the high-resolution crystal structures. Both the additive and Drude models populate regions consistent with the protein crystal structures, with the Drude model sampling a slightly wider range of  $\phi$ ,  $\psi$  space. In the helical region, the minimum in the Drude model is broader than with the additive

model, and there is additional sampling in the region of -120, 15. Both models have distributions similar to the survey results in the sheet regions, though C36 shows more well-defined minima with the Drude model exhibiting a broad, low energy region from  $\phi = -135$  to  $-60^\circ$  for  $\psi \sim 150^\circ$ . The Drude model also populates more of the region between the sheet and helical regions in the range of  $\psi = 30$ – $100^\circ$ . Finally, the Drude model distribution is broader in the  $\alpha$ L region as compared to C36, consistent with the remainder of the surface, with the location of the minima in both force fields consistent with the survey data.

The differences, as well as similarities, in the  $\phi$ ,  $\psi$  distributions are interesting when considered in the context of CMAP. While both the C36 and Drude CMAP surfaces have undergone some empirical adjustments, the underlying energy surfaces are based on quantum mechanics, such that the overall landscape of the surfaces should be similar. Adjustments in the C36 CMAP, which was obtained at the LMP2/cc-VQZ level, included local optimization of the helical and sheet regions to reproduce subtle features observed in crystallographic survey data<sup>4</sup> followed by subsequent shifting of the helical region to decrease the tendency for the C22/CMAP model to over-populate that conformation, leading to C36.<sup>5</sup> With the Drude model, the overall sheet region was lowered, and the areas between the sheet and helical regions and from  $\phi = -90$  to  $-180$  and  $\psi = -60$  to  $105^\circ$  were raised. Furthermore, we note that the conformational properties of the  $\chi_1$  and  $\chi_2$  side chain dihedrals were optimized with the same target function. Accordingly, the additional flexibility in the Drude model may largely be attributed to the inclusion of electronic polarization in the model. While further work is needed to understand this phenomenon, the result appears to be associated with the variability of the molecular dipoles in the Drude model, as detailed below.

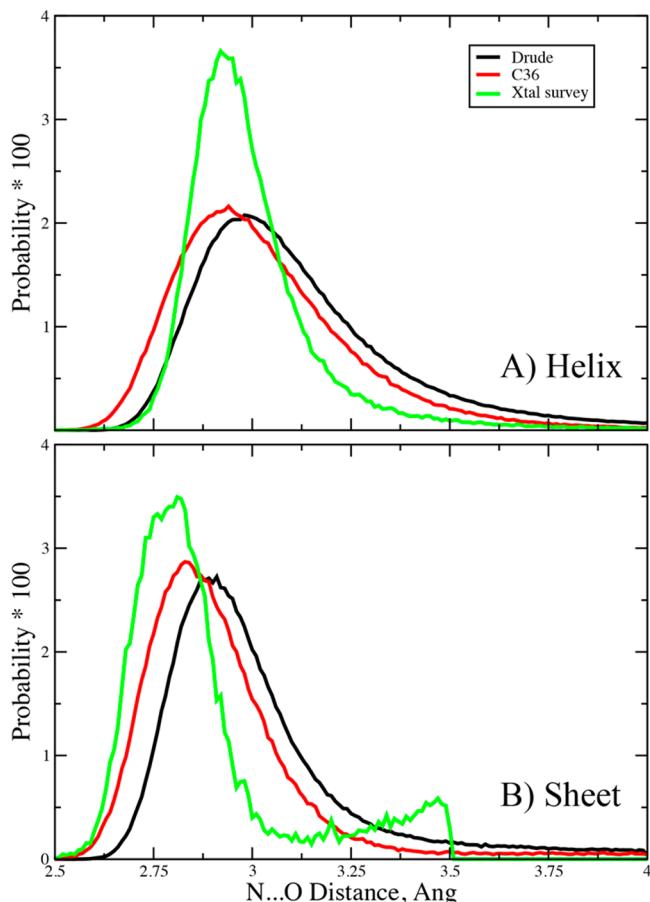
**NMR Analysis.** The above analysis largely involved the comparison of simulations of proteins in solution with



**Figure 4.** Helical properties in the dimeric coiled coil (1U01) simulations. (A) RMS differences of the individual helix C $\alpha$  atoms following alignment to themselves (e.g., Helix A vs Helix A) and following alignment based on the other helix in the dimer (e.g., Helix B vs Helix A). (B) Interhelical angle vectors defining the helical axes were calculated using the nonterminal residue C $\alpha$  atoms following Chothia et al.<sup>131</sup> NMR data for the 20 models generated in the published structural study are represented as horizontal lines.

experimental crystallographic data. To more carefully evaluate the behavior of the Drude model with respect to solution conditions, we calculated nuclear magnetic resonance (NMR) data, such as chemical shifts and  $S^2$  order parameters, for ubiquitin (1UBQ), protein GB3 (1P7E), and cold shock protein A (1MJC). This analysis builds upon our recent comparison of the C36 force field with NMR data.<sup>63,120–124</sup> A summary of the chemical shift results for six different nuclei for the three proteins is presented in Table 8; results for the three proteins are shown in Table S1 of the Supporting Information. The overall comparison indicates that the Drude model is in poorer agreement with experimental results than the highly optimized C36 model, consistent with the RMS difference analysis with respect to the crystal structures.

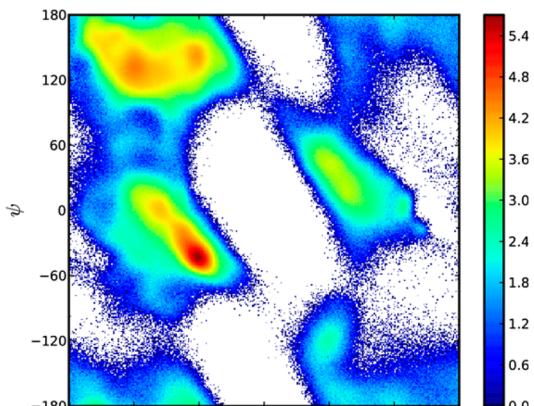
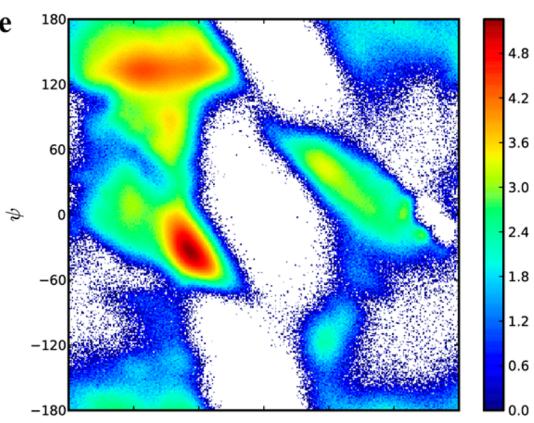
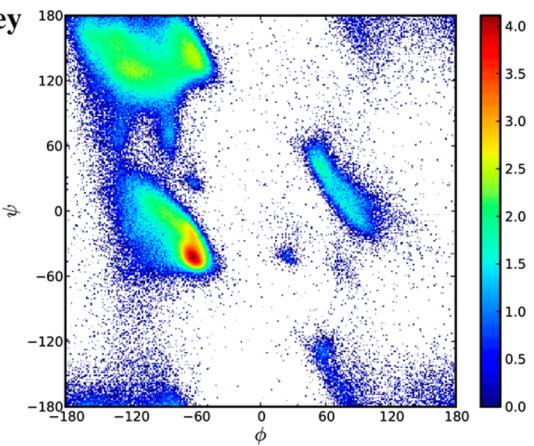
Additional NMR analysis involved the calculation of peptide backbone N–H order parameters,  $S^2$ , for the three proteins, with the results presented in Figure 7. The results in Figure 9 for C36 are from our previous study.<sup>63</sup> For all three proteins the additive and Drude FFs are generally in similar agreement with the experimental data. Notably, the Drude model does not systematically underestimate the  $S^2$  values as could have been expected given the enhanced backbone flexibility reflected in the above analyses. Moreover, in select cases, such as residues



**Figure 5.** Peptide N...O distance probability distributions for the (A) helical and (B) sheet secondary structures from Drude (black) and C36 (red) MD simulations of the proteins 1EJG, 3VQF, 3ZZP, and 4IEJ and from a survey of high-resolution crystal (xtal) structures in the Protein Databank. Probabilities were normalized to 1 and the values multiplied by 100 for clarity.

39 to 43 of ubiquitin and 8 to 11 and 73 in protein GB3, the Drude model gives significantly larger  $S^2$  values as compared to C36, with the Drude model being in better agreement with experimental results. These results suggest that, while the peptide backbone is populating a wider region of  $\phi$ ,  $\psi$  space, the range of conformations sampled is not significantly overestimated and, in some cases, may even be more realistic than with the additive force field. Such behavior is also consistent with the proteins in the Drude force field visiting a wider range of conformations in the MD simulations, though remaining stably folded on the time scale of the presented simulations.

**Protein Dipole Moment Analysis.** During the development of the electrostatic aspect of the force field, the partial atomic charges were put into groups of total unit charge. For example, the side chains, from the C $\beta$  atom onward, have partial atomic charges that sum to -1, 0, or 1. Similarly, the peptide bonds along with the C $\alpha$ /H $\alpha$  atoms have a total charge of 0. While this simplifying constraint was used to facilitate the transfer of the charges from the model compounds to the full biopolymer, it also has the advantage that it allows for the calculation of the dipole moments of the different functional groups in the MD simulations. Accordingly, analysis was performed on the dipole moments of the peptide bonds in the GB1 peptide, for the  $\beta$ -sheets and  $\alpha$ -helices in ubiquitin and the

**A) C36****B) Drude****C) Survey**

**Figure 6.** Overall  $\phi$ ,  $\psi$  distributions for the (A) C36 and (B) Drude model and a distribution from a survey of high-resolution crystal structures. Simulated data include results from 1EJG, 1P7E, 1MJC, 1UBQ, 3ZZP, 4IEJ, 3VQF, 13SL, 1IFC, and 1BYI.

Trp residues in lysozyme, which contains six such residues. For GB1 and lysozyme, the data are presented as time series of running averages over 10 ns, while the ubiquitin results are presented as probability distributions.

Analysis of the dipole moments in the backbone of GB1 shows large variability in the Drude model as compared to C36, as should be expected (Figure 8a). Variations in the additive model can only be associated with changes in the internal geometry of the peptide bond, with the occurrence of values around 4.6 D associated with the  $\alpha L$  conformation. It is particularly noteworthy that the values are systematically higher with the Drude model—even though the dipole moments in the additive force field are systematically overestimated in order to account for the polar environment in a mean-field manner.

**Table 8. Combined RMS Differences between Calculated and Experimental Chemical Shifts for Ubiquitin (1UBQ), Protein GB3 (1P7E) and Cold Shock Protein A (1MJC) with the C36 Additive and Drude Force Fields<sup>a</sup>**

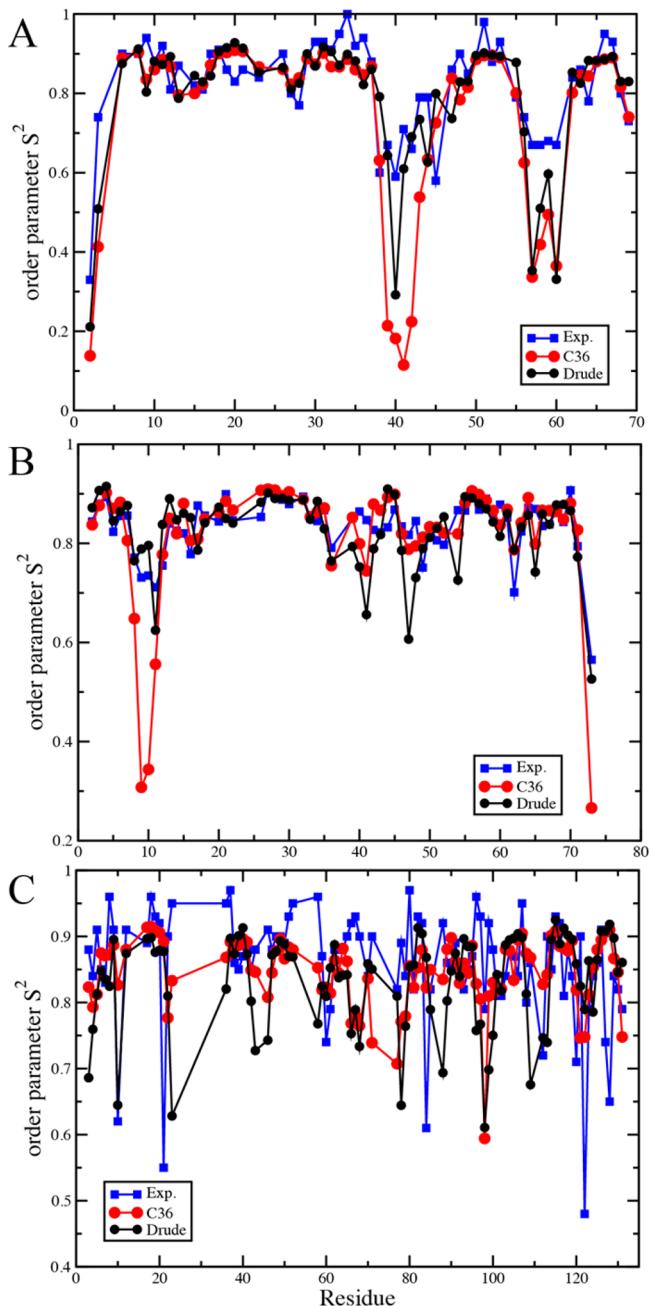
nucleus	# of nuclei	RMSD (ppm)	
		C36	Drude
N	192	2.50	3.21
C $\alpha$	201	0.76	1.03
C $\beta$	180	1.01	1.21
C'	141	0.95	0.98
H	193	0.41	0.47
H $\alpha$	201	0.23	0.30
all C	522	0.90	1.08
all H	394	0.33	0.39

<sup>a</sup>Number of each nuclei used in the RMS difference calculations is listed.

Apparently, the results from the Drude model indicate that this overestimation of the dipoles in the additive model is insufficient to account for the relevant environments encountered by the amino acids.

The peptide bond dipole moments were also analyzed with respect to secondary structure from the 1UBQ simulations. Shown in Figure 8b are probability distributions of the dipole moments in both helices and sheets for both the additive and polarizable FFs. The polarizable sheet distribution peaks in the vicinity of 5.3 D, consistent with the results for the GB1 hairpin. In the helices, the polarizable model predicts lower peptide bond dipole moments, with a maximum in the vicinity of 4.6 D. However, with the additive model, the maxima for the dipole moments are in the region of 3.7 D for both the sheets and helices, with the narrow distributions indicative of the fixed charge nature of the model. We note that the dipole moment in the additive model is close to the value of 4.12 D of the additive NMA model,<sup>3</sup> while that in the polarizable model is 3.72 D.<sup>96</sup> Clearly, the present results indicate that the impact of the environment on the peptide backbone is significant, leading to significant enhancements in the local peptide bond dipole moments. Notably, these enhancements are larger than the inherent dipole moment in the nonpolarizable additive force field, even though these moieties were “overpolarized” by design in the parametrization of this force field. The present observation is also consistent with previous studies based on both empirical and QM methods showing the protein and aqueous environments to alter the partial atomic charges in proteins.<sup>26,125,126</sup>

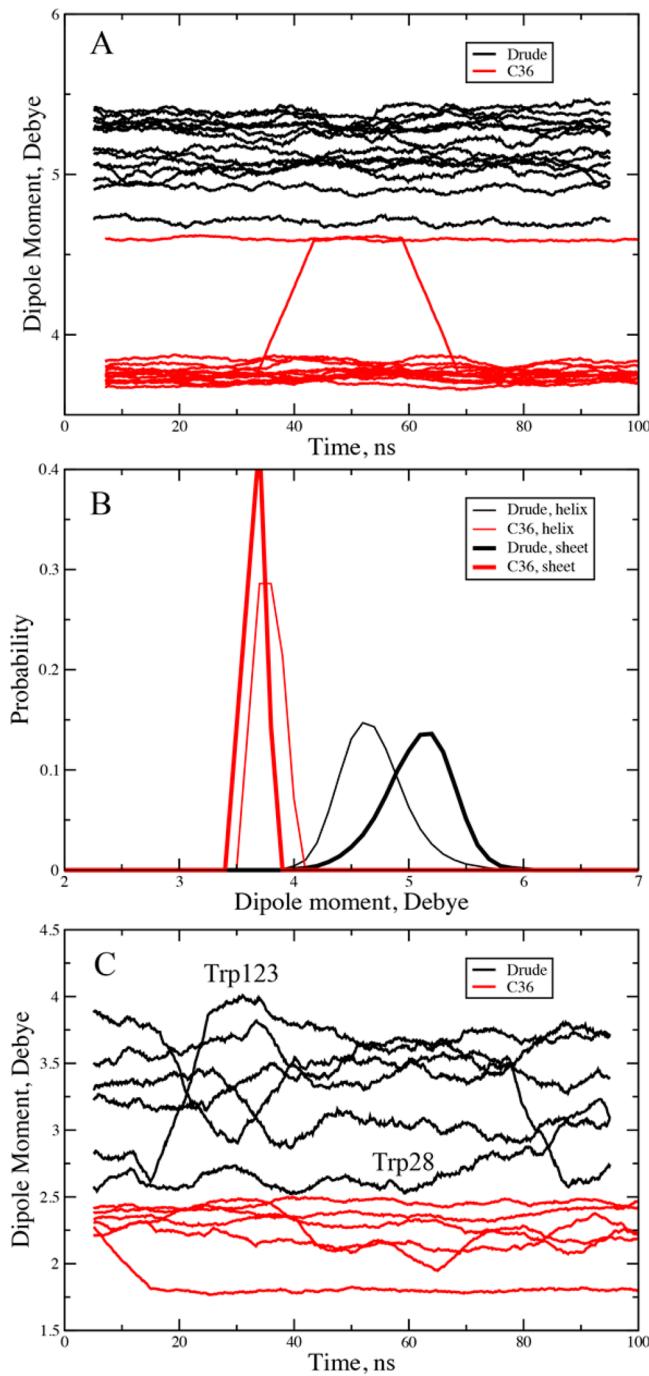
The increase in the peptide backbone dipole moments for both sheets and helices may be attributed to the hydrogen bonding interactions with surrounding peptide bonds in the secondary structures. However, the enhanced dipole moment of the sheets over the helices is somewhat unexpected. This appears to be due to the peptide dipoles pointing in opposite directions in the extended, sheet conformations, such that electrostatic interactions of peptide bond  $i$  with the adjacent  $i - 1$  and  $i + 1$  peptide bonds leads to enhancement of the peptide backbone dipole moment. This is similar to the phenomenon occurring in the Drude-1 model that leads to the stabilization of the C5 conformation in (Ala)<sub>5</sub> discussed above (Figure 2). In helices, the  $i - 1$  and  $i + 1$  peptide bonds are approximately parallel to that of the  $i$  peptide bond so the adjacent dipole moments do not enhance each other to as great an extent.



**Figure 7.** NMR N–H backbone order parameters,  $S_2$ , from calculations using both the C36 additive and Drude force fields for (A) ubiquitin (1UBQ), (B) protein GB3 (1MJC), and (C) cold shock protein A (1P7E).

Further studies into this observation, including the impact of changes in the electrostatic model, are warranted.

To understand the impact of the explicit inclusion of polarizability on the side chains, we focus on the six tryptophan residues in lysozyme. Presented in Figure 8c are the dipole moments as a function of time for those residues. With the additive model, the dipole moments are in the vicinity of 2.3 D, with small variations associated with changes in intramolecular geometry, including rotations about the  $\chi_2$  dihedral. In contrast, the dipole moments in the Drude model are significantly higher and fluctuate over a much wider range of values (from 2.5 to 4 D). Moreover, there are significant variations in the dipoles during the simulations, with the dipole

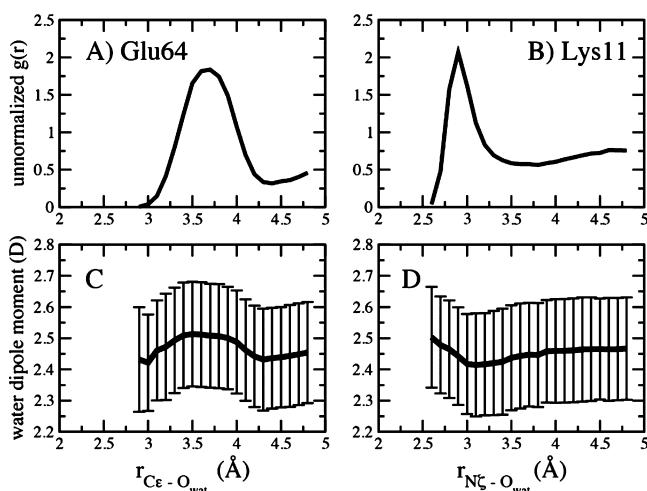


**Figure 8.** Dipole moment analysis (Drude, black; C36, red). (A) Peptide backbone dipoles as a function of time from the GB1 hairpin simulations. (B) Probability distribution of the peptide backbone dipole moments in the helices (thin lines) and  $\beta$ -sheets (thick lines) in ubiquitin (1UPQ). (C) Tryptophan side chain dipole moments as a function of time from the lysozyme (13SL) Drude B simulation, with the time series for residue Trp28 and Trp123 labeled. Time series are running averages over 100 0.1 ns windows.

moments of selected residues varying by more than 1 D during the simulations; analysis of individual snapshots from the simulations indicates variations from 1 to over 5 D associated with local high energy states accessible to local regions of structures during MD simulations (not shown). The tryptophan with the smallest dipole in the Drude simulation is Trp28, which is occluded from the solvent, being in contact

with the side chain Trp108 (Figure S5a, Supporting Information). The tryptophan with the largest variation is Trp123, whose dipole moment ranges from 2.6 to over 4 D during the simulation. This residue is solvent exposed and located adjacent to two helices (Figure S5b, Supporting Information), allowing it to sample different environments during the simulations, leading the variation in the dipole. As with the peptide backbone, the Drude model predicts the dipole moments of the Trp side chains to be significantly larger than in the additive force field with those dipoles being significantly different as a function of the local environment. The gas phase dipole moment of methylindole in the Drude model is 1.97,<sup>47</sup> compared to a value of 2.15 with the additive model,<sup>127</sup> indicating the significant polarization of the side chain that is occurring in the heterogeneous environment of lysozyme in aqueous solution.

**Water Dipole Moment Analysis.** Previous studies have shown the dipole moments of water to be perturbed in the vicinity of the protein in polarizable force field simulations.<sup>24,26</sup> Accordingly, we undertook analysis of the distribution of water molecules around the charged moieties of Glu64 and Lys11 of ubiquitin along with the change in the dipole moment of water as a function of distance (Figure 9). In both cases, it may be



**Figure 9.** “Un-normalized” radial distribution function,  $g(r)$ , and average dipole moments for water around (A and C) Glu64 and (B and D) Lys11 of ubiquitin. Distances based on the Glu  $C\epsilon$  or Lys  $N\zeta$  and the water oxygen, and the error bars represent the RMS fluctuations.

seen that there is a perturbation of the dipole of water in the vicinity of the charged groups. With Glu64 at short distances, the water dipole is decreased, and there is an increase of the dipole at the first maximum of the RDF followed by a decrease at the first minimum in the RDF. With Lys11, the dipole moment of water is enhanced at the contact distance and has a minimum just beyond the peak in the  $g(r)$  followed by an increase to the bulk value of 2.45 D at longer distances. The magnitude of the changes in the dipole moments appear to be slightly smaller than previously reported,<sup>24,26</sup> which may be due to scaling of the polarizable in the SWM4-NDP model by 0.7 as required to reproduce the dielectric constant of the model as well as other properties.<sup>38</sup>

**Practical Considerations.** MD simulations with a polarizable model display an enhanced sensitivity to initial conditions and are generally less robustly stable than

simulations carried out with an additive force field and may display polarization catastrophes that will lead to crashes. For this reason, MD simulations with a polarizable model require careful equilibration. Accordingly, the initial setup and equilibration of a simulation is best performed using an additive FF such as CHARMM36. This includes the initial solvation, minimization, and dynamics in the presence of restraints on the protein, thereby allowing the solvent to relax around the protein, with the dynamics being performed in the NVT ensemble. Additional equilibration should be performed using the additive model in the NPT ensemble with the protein structure allowed to move freely or subjected to weak harmonic restraints. A total of 100 ps of equilibration time may be sufficient, although extensive equilibration of 1 ns or more is recommended. The equilibrated system is then converted to the Drude polarizable model, including the ions and water, and subjected to an equilibration protocol similar to that initially performed using the additive model. As discussed above, protein simulations with the Drude model can typically be performed with a 1 fs integration time step, though a shorter time step may be required during the initial stage of equilibration as well as for highly charged systems.

To facilitate performing MD simulations with the Drude model, a new module, “Drude Prepper,” has been added to the CHARMM-GUI.<sup>90</sup> This module allows for CHARMM additive model protein structure file (PSF) and an equilibrated coordinate files to be uploaded and then converted to Drude formatted files, including all nomenclature changes. In addition, input scripts to perform MD simulations using either CHARMM or NAMD are produced. It is anticipated that such a utility should greatly facilitate studies of proteins based on the classical Drude polarizable force field.

## SUMMARY

Presented is a polarizable empirical force field based on the classical Drude oscillator for the modeling and simulation of proteins. In a first round, the parameters optimized from a large body of experimental and QM data on small model compounds representative of the all relevant protein functionalities were combined to yield a force field for the polymer. However, direct transfer of the parameters was not sufficient, requiring additional parameter optimization using larger model compounds. Most notable was the transfer of electrostatic parameters based on NMA to the polypeptide backbone. As this initial model was clearly insufficient, additional target data on the alanine dipeptide and longer polypeptides, including interactions with water, relative energies of different conformations, dipole moments, and molecular polarizabilities, were used to obtain the final (Drude-3) model. In addition, final fine-tuning of the Drude-3 model required additional adjustment to the relative potential energies of different regions of  $\phi$ ,  $\psi$  conformational space using the CMAP term in the energy function. The resulting model was shown to yield stable explicit solvent MD simulations on the 100 ns time scale for a collection of peptides and proteins. Notably, the Drude model displays more conformational flexibility than the C36 additive FF as estimated from RMS differences with respect to crystal structures and the distribution of  $\phi$ ,  $\psi$  conformations sampled in MD simulations. However, analysis of backbone N–H order parameters indicates that backbone fluctuations with the Drude model are not inconsistent with NMR experiments and, in specific cases, may even represent some improvement over C36.

Both the additive C36 and polarizable Drude models targeted similar, though not identical, QM data for development of both the CMAP energy term that defines the  $\phi$ ,  $\psi$  potential energy surface and the dihedral parameters dictating the  $\chi_1$  and  $\chi_2$  side chain torsional energetics followed by additional empirical optimization of these terms. Thus, as the underlying potential energy surfaces for backbone and side chain sampling may be considered similar, the implication is that the systematically enhanced conformational flexibility of the Drude model must be associated with the electrostatic model that includes explicit induced polarization. Analysis of molecular dipoles in both the additive and Drude models shows that the polarizable model gives systematically larger dipole moments for the studied moieties as compared to the nonpolarizable additive model, with those dipole moments showing significant variation during MD simulations. It therefore appears that the variations of the electronic structure do impact the dynamics of the system and the microscopic forces dictating the structural and dynamical properties of proteins. Interestingly, we note that the polarizabilities for many of the moieties in the polypeptides were scaled down relative the QM data by factors ranging from 1.0 with the aliphatics, to 0.6 with the sulfur containing groups. Indeed, the molecular polarizability of the final backbone model is systematically lower than QM estimate for the alanine dipeptide (Table 4, note that the QM values are scaled by 0.85). Thus, it seems unlikely that the current model is inherently overpolarized, though this possibility requires further study.

Overall, the current model should be considered a first generation polarizable force field that can be used for MD simulations on the order of hundreds of nanoseconds. Comparison with the nonpolarizable C36 additive force field, which has had the advantage of over 20 years of testing and additional optimization since its initial completion, suggests that further improvements in the model ought to be possible. Accordingly, we anticipate that an improved second generation model can and will be produced. Nevertheless, we are confident that the current model does provide an accurate and improved picture of the structure and function of polypeptides and proteins and that it can serve as an important computational tool for the biophysical and chemical scientific communities. We also note the availability of enhancements in the model that yield more accurate treatment of cation ion- $\pi$  interactions involving aromatic residues,<sup>128</sup> of an improved, though computationally more expensive Drude water model,<sup>39</sup> and of lipid Drude parameters<sup>53</sup> and the anticipated availability of parameters for nucleic acids (Savelyev and MacKerell, work in progress) and carbohydrates<sup>54</sup> as well as tools for the optimization of small molecule parameters.<sup>129</sup> These enhancements and capabilities will allow for computational studies of heterogeneous systems using a fully polarizable force field.

The final residue topology information and parameters in CHARMM format may be downloaded from the MacKerell Lab Web site at [http://mackerell.umaryland.edu/MacKerell\\_Lab.html](http://mackerell.umaryland.edu/MacKerell_Lab.html).

## ASSOCIATED CONTENT

### Supporting Information

Included are figures of the interactions of the alanine dipeptide with water, cartoon images of the studied peptides and proteins, RMS difference plots, Ramachandran  $\phi$ ,  $\psi$  2D surfaces, images of selected tryptophans in lysozyme, comparison of calculated

and experimental NMR chemical shift data, and the topology and parameter information for the presented Drude force field. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## AUTHOR INFORMATION

### Corresponding Authors

\*E-mail: roux@uchicago.edu.

\*E-mail: alex@outerbanks.umaryland.edu.

### Present Address

<sup>†</sup>Department of Pharmaceutical Sciences, Western University of Health Sciences, 309 E Second Street, Pomona, CA 91766

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

Financial support from the NIH (GM072558) and computational support from the University of Maryland Computer-Aided Drug Design Center and the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number OCI-1053575, are acknowledged. J.H. is supported by a Swiss National Science Foundation Fellowship (PBBSP2\_144301). Y.L. was supported by an Early Science Project Fellowship from the Argonne Leadership Computing Facility (ALCF) at Argonne National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under contract DE-AC02-06CH11357. A special thanks to Jim Phillips, David Hardy, Klaus Schulten, Wei Jiang, and Lei Huang for their efforts in implementing the Drude model in NAMD, which was critical for this work, and to all the members of the Roux and MacKerell groups who contributed to the development of the Drude model over the years.

## REFERENCES

- Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
- Wang, Z. X.; Zhang, W.; Wu, C.; Lei, H.; Cieplak, P.; Duan, Y. Strike a balance: optimization of backbone torsion parameters of AMBER polarizable force field for simulations of proteins and peptides. *J. Comput. Chem.* **2006**, *27*, 781–790.
- MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, I. W. E.; Roux, B.; Schlenkrich, M.; Smith, J.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
- MacKerell, A. D., Jr.; Feig, M.; Brooks, C. L., III. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **2004**, *25*, 1400–1415.
- Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell, A. D., Jr. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone  $\varphi$ ,  $\psi$  and side-chain  $\chi_1$  and  $\chi_2$  dihedral angles. *J. Chem. Theory Comput.* **2012**, *8*, 3257–3273.
- Oostenbrink, C.; Soares, T. A.; van der Vegt, N. F. A.; van Gunsteren, W. F. Validation of the 53A6 GROMOS force field. *Eur. Biophys. J.* **2005**, *34*, 273–284.

- (7) Jorgensen, W. L.; Tirado-Rives, J. The OPLS Potential Function for Proteins. Energy Minimizations for Crystals of Cyclic Peptides and Crambin. *J. Am. Chem. Soc.* **1988**, *110*, 1657–1666.
- (8) Kaminski, G.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. *J. Phys. Chem. B* **2001**, *105*, 6474–6487.
- (9) Ponder, J. W.; Case, D. A. Force Fields for Protein Simulations. *Adv. Protein. Chem.* **2003**, *66*, 27–85.
- (10) MacKerell, A. D., Jr. Empirical Force Fields for Biological Macromolecules: Overview and Issues. *J. Comput. Chem.* **2004**, *25*, 1584–1604.
- (11) Warshel, A.; Levitt, M. Theoretical Studies of Enzymic Reactions: Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme. *J. Mol. Biol.* **1976**, *103*, 227–249.
- (12) Russell, S. T.; Warshel, A. Calculations of Electrostatic Energies in Proteins: The Energetics of Ionized Groups in Bovine Pancreatic Trypsin Inhibitor. *J. Mol. Biol.* **1985**, *185*, 389–404.
- (13) van Belle, D.; Couplet, I.; Prevost, M.; Wodak, S. J. Calculations of Electrostatic Properties in Proteins - Analysis of Contributions from Induced Protein Dipoles. *J. Mol. Biol.* **1987**, *198*, 721–735.
- (14) Halgren, T. A.; Damm, W. Polarizable force fields. *Curr. Opin. Struct. Biol.* **2001**, *11*, 236–242.
- (15) Rick, S. W.; Stuart, S. J. Potentials and Algorithms for Incorporating Polarizability in Computer Simulations. *Rev. Comp. Chem.* **2002**, *18*, 89–146.
- (16) Warshel, A.; Kato, M.; Pisliakov, A. V. Polarizable Force Fields: History, Test Cases, and Prospects. *J. Chem. Theory Comput.* **2007**, *3*, 2034–2045.
- (17) Lopes, P. E. M.; Roux, B.; MacKerell, A. D., Jr. Molecular modeling and dynamics studies with explicit inclusion of electronic polarizability: theory and applications. *Theor. Chem. Acc.* **2009**, *124*, 11–28.
- (18) Cieplak, P.; Dupradeau, F.-Y.; Duan, Y.; Wang, J. Polarization effects in molecular mechanical force fields. *J. Phys.: Condens. Matter* **2009**, *21*, 333102.
- (19) Banks, J. L.; Kaminski, G. A.; Zhou, R.; Mainz, D. T.; Berne, B. J.; Friesner, R. A. Parametrizing a Polarizable Force Field from Ab Initio Data. I. The Fluctuating Point Charge Model. *J. Chem. Phys.* **1999**, *110*, 741–754.
- (20) Stern, H. A.; Kaminski, G. A.; Banks, J. L.; Zhou, R.; Berne, B. J.; Friesner, R. A. Fluctuating Charge, Polarizable Dipole, and Combined Models: Parameterization from ab Initio Quantum Chemistry. *J. Phys. Chem. B* **1999**, *103*, 4730–4737.
- (21) Stern, H. A.; Rittner, F.; Berne, B. J.; Friesner, R. A. Combined fluctuating charge and polarizable dipole models: Application to a five-site water potential function. *J. Chem. Phys.* **2001**, *115*, 2237–2251.
- (22) Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A.; Cao, Y. X.; Murphy, R. B.; Zhou, R.; Halgren, T. A. Development of a Polarizable Force Field for Proteins via Ab Initio Quantum Chemistry: First Generation Model and Gas Phase Tests. *J. Comput. Chem.* **2002**, *23*, 1515–1531.
- (23) Harder, E.; Kim, B.; Friesner, R. A.; Berne, B. J. Efficient Simulation Method for Polarizable Protein Force Fields: Application to the Simulation of BPTI in Liquid Water. *J. Chem. Theory Comput.* **2005**, *1*, 169–180.
- (24) Kim, B.; Young, T.; Harder, E.; Friesner, R. A.; Berne, B. J. Structure and Dynamics of the Solvation of Bovine Pancreatic Trypsin Inhibitor in Explicit Water: A Comparative Study of the Effects of Solvent and Protein Polarizability. *J. Phys. Chem. B* **2005**, *109*, 16529–16538.
- (25) Patel, S.; Brooks, C. L., III. CHARMM fluctuating charge force field for proteins: I parameterization and application to bulk organic liquid simulations. *J. Comput. Chem.* **2004**, *25*, 1–15.
- (26) Patel, S.; MacKerell, A. D., Jr.; Brooks, C. L., III. CHARMM fluctuating charge force field for proteins: II Protein/solvent properties from molecular dynamics simulations using a nonadditive electrostatic model. *J. Comput. Chem.* **2004**, *25*, 1504–1514.
- (27) Ren, P.; Ponder, J. W. Consistent Treatment of Inter- and Intramolecular Polarization in Molecular Mechanics Calculations. *J. Comput. Chem.* **2002**, *23*, 1497–1506.
- (28) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A., Jr.; Head-Gordon, M.; Clark, G. N.; Johnson, M. E.; Head-Gordon, T. Current status of the AMOEBA polarizable force field. *J. Phys. Chem. B* **2010**, *114*, 2549–2564.
- (29) Jiao, D.; Golubkov, P. A.; Darden, T. A.; Ren, P. Calculation of protein-ligand binding free energy by using a polarizable potential. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 6290–6295.
- (30) Jiao, D.; Zhang, J.; Duke, R. E.; Li, G.; Schnieders, M. J.; Ren, P. Trypsin-ligand binding free energies from explicit and implicit solvent simulations with polarizable potential. *J. Comput. Chem.* **2009**, *30*, 1701–1711.
- (31) Shi, Y.; Zhu, C. Z.; Martin, S. F.; Ren, P. Probing the Effect of Conformational Constraint on Phosphorylated Ligand Binding to an SH2 Domain Using Polarizable Force Field Simulations. *J. Phys. Chem. B* **2012**, *116*, 1716–1727.
- (32) Zhang, J.; Yang, W.; Piquemal, J. P.; Ren, P. Modeling Structural Coordination and Ligand Binding in Zinc Proteins with a Polarizable Potential. *J. Chem. Theory Comput.* **2012**, *8*, 1314–1324.
- (33) Schultheis, V.; Reichold, R.; Schropp, B.; Tavan, P. A Polarizable Force Field for Computing the Infrared Spectra of the Polypeptide Backbone. *J. Phys. Chem. B* **2008**, *112*, 12217–12230.
- (34) Click, T. H.; Kaminski, G. A. Reproducing basic pKa values for turkey ovomucoid third domain using a polarizable force field. *J. Phys. Chem. B* **2009**, *113*, 7844–7850.
- (35) Zhong, Y.; Patel, S. Binding structures of tri-N-acetyl-beta-glucosamine in hen egg white lysozyme using molecular dynamics with a polarizable force field. *J. Comput. Chem.* **2013**, *34*, 163–174.
- (36) Shi, Y.; Xia, Z.; Zhang, J.; Best, R.; Wu, C.; Ponder, J. W.; Ren, P. Polarizable Atomic Multipole-Based AMOEBA Force Field for Proteins. *J. Chem. Theory Comput.* **2013**, *9*, 4046–4064.
- (37) Lamoureux, G.; MacKerell, A. D., Jr.; Roux, B. A simple polarizable model of water based on classical Drude oscillators. *J. Chem. Phys.* **2003**, *119*, 5185–5197.
- (38) Lamoureux, G.; Harder, E.; Vorobyov, I. V.; Roux, B.; MacKerell, A. D., Jr. A polarizable model of water for molecular dynamics simulations of biomolecules. *Chem. Phys. Lett.* **2006**, *418*, 245–249.
- (39) Yu, W.; Lopes, P. E.; Roux, B.; MacKerell, A. D., Jr. Six-site polarizable model of water based on the classical Drude oscillator. *J. Chem. Phys.* **2013**, *138*, 034508.
- (40) Vorobyov, I.; Anisimov, V. M.; Greene, S.; Venable, R. M.; Moser, A.; Pastor, R. W.; MacKerell, A. D., Jr. Additive and Classical Drude Polarizable Force Fields for Linear and Cyclic Ethers. *J. Chem. Theory Comput.* **2007**, *3*, 1120–1133.
- (41) Harder, E.; Anisimov, V. M.; Whitfield, T.; MacKerell, A. D., Jr.; Roux, B. Understanding the Dielectric Properties of Liquid Amides from a Polarizable Force Field. *J. Phys. Chem. B* **2008**, *112*, 3509–3521.
- (42) Lopes, P. E. M.; Lamoureux, G.; Roux, B.; MacKerell, A. D., Jr. Polarizable Empirical Force Field for Aromatic Compounds Based on the Classical Drude Oscillator. *J. Phys. Chem. B* **2007**, *111*, 2873–2885.
- (43) Anisimov, V. M.; Vorobyov, I. V.; Roux, B.; MacKerell, A. D., Jr. Polarizable empirical force field for the primary and secondary alcohol series based on the classical Drude model. *J. Chem. Theory Comput.* **2007**, *3*, 1927–1946.
- (44) Harder, E.; Anisimov, V. M.; Vorobyov, I. V.; Lopes, P. E. M.; Noskov, S.; MacKerell, A. D., Jr.; Roux, B. Atomic Level Anisotropy in the Electrostatic Modeling of Lone Pairs for a Polarizable Force Field based on the Classical Drude Oscillator. *J. Chem. Theory Comput.* **2006**, *2*, 1587–1597.
- (45) Zhu, X.; MacKerell, A. D., Jr. Polarizable empirical force field for sulfur-containing compounds based on the classical Drude oscillator model. *J. Comput. Chem.* **2010**, *31*, 2330–2341.

- (46) Baker, C. M.; MacKerell, A. D., Jr. Polarizability rescaling and atom-based Thole scaling in the CHARMM Drude polarizable force field for ethers. *J. Mol. Model.* **2010**, *16*, 567–576.
- (47) Lopes, P. E. M.; Lamoureux, G.; MacKerell, A. D., Jr. Polarizable Empirical Force Field for Nitrogen-containing Heteroaromatic Compounds Based on the Classical Drude Oscillator. *J. Comput. Chem.* **2009**, *30*, 1821–1838.
- (48) Baker, C. M.; Anisimov, V. M.; MacKerell, A. D., Jr. Development of CHARMM Polarizable Force Field for Nucleic Acid Bases Based on the Classical Drude Oscillator Model. *J. Phys. Chem. B* **2011**, *115*, 580–596.
- (49) Yu, H.; Whitfield, T. W.; Harder, E.; Lamoureux, G.; Vorobyov, I.; Anisimov, V. M.; MacKerell, A. D., Jr.; Roux, B. Simulating Monovalent and Divalent Ions in Aqueous Solution Using a Drude Polarizable Force Field. *J. Chem. Theory Comput.* **2010**, *6*, 774–786.
- (50) Luo, Y.; Jiang, W.; Yu, H.; MacKerell, A. D., Jr.; Roux, B. Simulation study of ion pairing in concentrated aqueous salt solutions with a polarizable force field. *Faraday Discuss.* **2013**, *160*, 135–149.
- (51) Anisimov, V. M.; Lamoureux, G.; Vorobyov, I. V.; Huang, N.; Roux, B.; MacKerell, A. D., Jr. Determination of Electrostatic Parameters for a Polarizable Force Field Based on the Classical Drude Oscillator. *J. Chem. Theory Comput.* **2005**, *1*, 153–168.
- (52) Harder, E.; MacKerell, A. D., Jr.; Roux, B. Electronic Polarization Effects and the Membrane Dipole Potential. *J. Am. Chem. Soc.* **2009**, *131*, 2760–2761.
- (53) Chowdhary, J.; Harder, E.; Lopes, P. E.; Huang, L.; MacKerell, A. D., Jr.; Roux, B. A Polarizable Force Field of Dipalmitoylphosphatidylcholine Based on the Classical Drude Model for Molecular Dynamics Simulations of Lipids. *J. Phys. Chem. B* **2013**, *117*, 9142–9160.
- (54) He, X.; Lopes, P. E. M.; MacKerell, A. D., Jr. Polarizable Empirical Force Field for Acyclic Poly-Alcohols Based on the Classical Drude Oscillator. *Biopolymers* **2013**, *99*, 724–738.
- (55) Best, R. B.; Mittal, J. Balance between alpha and beta structures in ab initio protein folding. *J. Phys. Chem. B* **2010**, *114*, 8790–8798.
- (56) Best, R. B.; Mittal, J. Protein simulations with an optimized water model: cooperative helix formation and temperature-induced unfolded state collapse. *J. Phys. Chem. B* **2010**, *114*, 14916–14923.
- (57) Best, R. B.; Buchete, N. V.; Hummer, G. Are current molecular dynamics force fields too helical? *Biophys. J.* **2008**, *95*, L07–09.
- (58) Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y.; Wriggers, W. Atomic-level characterization of the structural dynamics of proteins. *Science* **2010**, *330*, 341–346.
- (59) Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. How robust are protein folding simulations with respect to force field parameterization? *Biophys. J.* **2011**, *100*, L47–49.
- (60) Lindorff-Larsen, K.; Maragakis, P.; Piana, S.; Eastwood, M. P.; Dror, R. O.; Shaw, D. E. Systematic validation of protein force fields against experimental data. *PLoS One* **2012**, *7*, e32131.
- (61) Piana, S.; Lindorff-Larsen, K.; Dirks, R. M.; Salmon, J. K.; Dror, R. O.; Shaw, D. E. Evaluating the effects of cutoffs and treatment of long-range electrostatics in protein folding simulations. *PLoS One* **2012**, *7*, e39918.
- (62) Best, R. B.; Mittal, J.; Feig, M.; MacKerell, A. D., Jr. Inclusion of many-body effects in the additive CHARMM protein CMAP potential results in enhanced cooperativity of alpha-helix and beta-hairpin formation. *Biophys. J.* **2012**, *103*, 1045–1051.
- (63) Huang, J.; MacKerell, A. D., Jr. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *J. Comput. Chem.* **2013**, *34*, 2135–2145.
- (64) van Marren, P. J.; van der Spoel, D. Molecular Dynamics Simulations of Water with Novel Shell-Model Potentials. *J. Phys. Chem. B* **2001**, *105*, 2618–2626.
- (65) Kunz, A. P.; van Gunsteren, W. F. Development of a nonlinear classical polarization model for liquid water and aqueous solutions: COS/D. *J. Phys. Chem. A* **2009**, *113*, 11570–11579.
- (66) Drude, P.; Millikan, R. A.; Mann, R. C. *The Theory of Optics*; Longmans, Green, and Co.: New York, 1902; p 588.
- (67) Waldman, M.; Gordon, R. G. Generalized electron gas-Drude model theory of intermolecular forces. *J. Chem. Phys.* **1979**, *71*, 1340–1352.
- (68) Applequist, J.; Carl, J. R.; Fung, K.-K. Atom dipole interaction model for molecular polarizability. Application to polyatomic molecules and determination of atom polarizabilities. *J. Am. Chem. Soc.* **1972**, *94*, 2952–2960.
- (69) Thole, B. T. Molecular polarizabilities calculated with a modified dipole interaction. *Chem. Phys.* **1981**, *59*, 341–350.
- (70) Lopes, P. E. M.; Harder, E.; Roux, B.; MacKerell, A. D., Jr., Formalisms for the explicit inclusion of electronic polarizability in molecular modeling and dynamics studies. In *Multi-scale Quantum Models for Biocatalysis: Modern Techniques and Applications*; York, D., Lee, T.-S., Eds.; Springer: New York, 2009.
- (71) Sprik, M.; Klein, M. L. A Polarizable Model for Water Using Distributed Charge Sites. *J. Chem. Phys.* **1988**, *89*, 7556–7560.
- (72) Martyna, G. J.; Tuckerman, M. E.; Tobias, D. J.; Klein, M. L. Explicit reversible integrators for extended systems dynamics. *Mol. Phys.* **1996**, *87*, 1117–1157.
- (73) Tuckerman, M. E.; Martyna, G. J. Understanding Modern Molecular Dynamics: Techniques and Applications. *J. Phys. Chem. B* **2000**, *104*, 159–178.
- (74) Lamoureux, G.; Roux, B. Modelling Induced Polarizability with Drude Oscillators: Theory and Molecular Dynamics Simulation Algorithm. *J. Chem. Phys.* **2003**, *119*, 5185–5197.
- (75) Jiang, W.; Hardy, D. J.; Phillips, J. C.; MacKerell, A. D., Jr.; Schulten, K.; Roux, B. High-performance scalable molecular dynamics simulations of a polarizable force field based on classical Drude oscillators in NAMD. *J. Phys. Chem. Lett.* **2011**, *2*, 87–92.
- (76) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*; Gaussian, Inc.: Wallingford, CT, 2009.
- (77) Shao, Y.; Fusti-Molnar, L.; Jung, Y.; Kussmann, J.; Ochsenfeld, C.; Brown, S. T.; Gilbert, A. T. B.; Slipchenko, L. V.; Levchenko, S. V.; O'Neill, D. P.; DiStasio, R. A., Jr.; Lochan, R. C.; Wang, T.; Beran, G. J. O.; Besley, N. A.; Herbert, J. M.; Lin, C. Y.; Voorhis, T. V.; Chien, S. H.; Sodt, A.; Steele, R. P.; Rassolov, V. A.; Maslen, P. E.; Korambath, P. P.; Adamson, R. D.; Austin, B.; Baker, J.; Byrd, E. F. C.; Dachselt, H.; Doerkens, R. J.; Dreuw, A.; Dunietz, B. D.; Dutoi, A. D.; Furlani, T. R.; Gwaltney, S. R.; Heyden, A.; Hirata, S.; Hsu, C.-P.; Kedziora, G.; Khalilulin, R. Z.; Klunzinger, P.; Lee, A. M.; Lee, M. S.; Liang, W.; Lotan, I.; Nair, N.; Peters, B.; Proynov, E. I.; Pieniazek, P. A.; Rhee, Y. M.; Ritchie, J.; Rosta, E.; Sherrill, C. D.; Simonett, A. C.; Subotnik, J. E.; Woodcock, H. L., III; Zhang, W.; Bell, A. T.; Chakraborty, A. K.; Chipman, D. M.; Keil, F. J.; Warshel, A.; Hehre, W. J.; Schaefer, H. F., III; Kong, J.; Krylov, A. I.; Gill, P. M. W.; Head-Gordon, M.; Gan, Z.; Zhao, Y.; Schultz, N. E.; Truhlar, D.; Epifanovsky, E.; Oana, M. Q. *Chem. 3.1*; Q-Chem, Inc.: Pittsburgh, PA, 2007.
- (78) Kong, J.; White, C. A.; Krylov, A. I.; Sherrill, C. D.; Adamson, R. D.; Furlani, T. R.; Lee, M. S.; Lee, A. M.; Gwaltney, S. R.; Adams, T. R.; Ochsenfeld, C.; Gilbert, A. T. B.; Kedziora, G. S.; Rassolov, V. A.; Maurice, D. R.; Nair, N.; Shao, Y.; Besley, N. A.; Maslen, P. E.; Dombroski, J. P.; Daschel, H.; Zhang, W.; Korambath, P. P.; Baker, J.; Byrd, E. F. C.; Voorhis, T. V.; Oumi, M.; Hirata, S.; Hsu, C.-P.; Ishikawa, N.; Florian, J.; Warshel, A.; Johnson, B. G.; Gill, P. M. W.; Head-Gordon, M.; Pople, J. A. Q-Chem 2.0: A high-performance ab

- initio electronic structure program. *J. Comput. Chem.* **2000**, *21*, 1532–1548.
- (79) Woon, D. E.; Dunning, T. H., Jr. Gaussian basis sets for use in correlated molecular calculations. III. The atoms aluminum through argon. *J. Chem. Phys.* **1993**, *98*, 1358–1371.
- (80) Møller, C.; Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev.* **1934**, *46*, 618–622.
- (81) Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Kocha, H.; Olsenc, J.; Wilson, A. K. Basis-set convergence in correlated calculations on Ne, N<sub>2</sub>, and H<sub>2</sub>O. *Chem. Phys. Lett.* **1998**, *286*, 243–252.
- (82) Brooks, B. R.; Brooks, C. L., III; MacKerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **2009**, *30*, 1545–1614.
- (83) MacKerell, A. D., Jr.; Brooks, B.; Brooks, C. L., III; Nilsson, L.; Roux, B.; Won, Y.; Karplus, M., CHARMM: The Energy Function and Its Parameterization with an Overview of the Program. In *Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P. A., Schaefer, H. F., III, Schreiner, P. R., Eds.; John Wiley & Sons: Chichester, U. K., 1998; Vol. 1, pp 271–277.
- (84) Brooks, B. R.; Brucoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.* **1983**, *4*, 187–217.
- (85) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford University Press: New York, 1989.
- (86) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.
- (87) Miyamoto, S.; Kollman, P. SETTLE: An Analytical Version of the SHAKE and RATTLE Algorithm for Rigid Water Models. *J. Comput. Chem.* **1992**, *13*, 952–962.
- (88) Berman, H. M.; Battistuz, T.; Bhat, T. N.; Bluhm, W. F.; Bourne, P. E.; Burkhardt, K.; Feng, Z.; Gilliland, G. L.; Iype, L.; Jain, S.; Fagan, P.; Marvin, J.; Padilla, D.; Ravichandran, V.; Schneider, B.; Thanki, N.; Weissig, H.; Westbrook, J. D.; Zardecki, C. The protein data bank. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2002**, *58*, 899–907.
- (89) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (90) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: a web-based graphical user interface for CHARMM. *J. Comput. Chem.* **2008**, *29*, 1859–1865.
- (91) Fukunishi, H.; Watanabe, O.; Takada, S. On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction. *J. Chem. Phys.* **2002**, *116*, 9058.
- (92) Graf, J.; Nguyen, P. H.; Stock, G.; Schwalbe, H. Structure and dynamics of the homologous series of alanine peptides: A joint molecular dynamics/NMR study. *J. Am. Chem. Soc.* **2007**, *129*, 1179–1189.
- (93) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. Equation of State Calculations by Fast Computing Machines. *J. Chem. Phys.* **1953**, *21*, 1087–1092.
- (94) Hansmann, U. H. E. Parallel tempering algorithm for conformational studies of biological molecules. *Chem. Phys. Lett.* **1997**, *281*, 140–150.
- (95) Sugita, Y.; Kitao, A.; Okamoto, Y. Multidimensional replica-exchange method for free-energy calculations. *J. Chem. Phys.* **2000**, *113*, 6042–6051.
- (96) Lin, B.; Lopes, P. E. M.; Roux, B.; MacKerell, A. D., Jr. Kirkwood-Buff Analysis of Aqueous N-Methylacetamide and Acetamide Solutions Modeled by the CHARMM Additive and Drude Polarizable Force Fields. *J. Chem. Phys.* **2013**, *139*, 084509.
- (97) Miller, K. J. Additivity Methods in Molecular Polarizability. *J. Am. Chem. Soc.* **1990**, *112*, 8533–8542.
- (98) Kirkpatrick, S.; Gelatt, C. D. J.; Vecchi, M. P. Optimization by Simulated Annealing. *Science* **1983**, *220*, 671–680.
- (99) Zhu, X.; Lopes, P. E.; Shim, J.; MacKerell, A. D., Jr. Intrinsic energy landscapes of amino acid side-chains. *J. Chem. Inf. Model.* **2012**, *S2*, 1559–1572.
- (100) Guvench, O.; MacKerell, A. D., Jr. Automated conformational energy fitting for force-field development. *J. Mol. Mod.* **2008**, *14*, 667–679.
- (101) Wang, L.; Friesner, R. A.; Berne, B. J. Replica exchange with solute scaling: a more efficient version of replica exchange with solute tempering (REST2). *J. Phys. Chem. B* **2011**, *115*, 9431–9438.
- (102) Liu, P.; Kim, B.; Friesner, R. A.; Berne, B. J. Replica exchange with solute tempering: A method for sampling biological systems in explicit water. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 13749–13754.
- (103) Bernard, D.; Coop, A.; MacKerell, A. D., Jr. 2D Conformationally Sampled Pharmacophore: A Ligand-Based Pharmacophore To Differentiate delta Opoid Agonists from Antagonists. *J. Am. Chem. Soc.* **2003**, *125*, 3101–3107.
- (104) Foloppe, N.; MacKerell, A. D., Jr. All-atom empirical force field for nucleic acids: 1) Parameter optimization based on small molecule and condensed phase macromolecular target data. *J. Comput. Chem.* **2000**, *21*, 86–104.
- (105) Oostenbrink, C.; Villa, A.; Mark, A. E.; van Gunsteren, W. F.; Biomolecular Force, A. Field Based on the Free Enthalpy of Hydration and Solvation: The GROMOS Force-Field Parameter Sets 53A5 and 53A6. *J. Comput. Chem.* **2004**, *25*, 1656–1676.
- (106) MacKerell, A. D., Jr.; Feig, M.; Brooks, C. L., III. Accurate treatment of protein backbone conformational energetics in empirical force fields. *J. Am. Chem. Soc.* **2004**, *126*, 698–699.
- (107) Shoemaker, K. R.; Kim, P. S.; Brems, D. N.; Marqusee, S.; York, E. J.; Chaiken, I. M.; Stewart, J. M.; Baldwin, R. L. Nature of the Charged-Group Effect on the Stability of the C-Peptide Helix. *Proc. Natl. Acad. Sci. U. S. A.* **1985**, *82*, 2349–2353.
- (108) Shoemaker, K. R.; Kim, P. S.; York, E. J.; Stewart, J. M.; Baldwin, R. L. Tests of the Helix Dipole Model for Stabilization of  $\alpha$ -Helices. *Nature* **1987**, *326*, 563–567.
- (109) Padmanabhan, S.; Marqusee, S.; Ridgeway, T.; Laue, T. M.; Baldwin, R. L. Relative Helix-Forming Tendencies of Nonpolar Amino Acids. *Nature* **1990**, *344*, 268–270.
- (110) Dunbrack, R. L.; Karplus, M. Backbone-Dependent Rotamer Library for Proteins. Application to Side-Chain Prediction. *J. Mol. Biol.* **1993**, *230*, 543–574.
- (111) Dunbrack, R. L., Jr.; Cohen, F. E. Bayesian statistical analysis of protein sidechain rotamer preferences. *Protein Sci.* **1997**, *6*, 1661–1681.
- (112) Dunbrack, R. L., Jr. Culledpdb: Non-redundant set of protein sidechains from the PDB. <http://www.fccc.edu/research/labs/dunbrack/culledpdb.html>.
- (113) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010**, *78*, 1950–1958.
- (114) Shim, J.; Zhu, X.; Best, R. B.; MacKerell, A. D., Jr. (Ala)(4)-X-(Ala)4 as a model system for the optimization of the chi1 and chi2 amino acid side-chain dihedral empirical force field parameters. *J. Comput. Chem.* **2013**, *34*, 593–603.
- (115) Blanco, F. J.; Rivas, G.; Serrano, L. A short linear peptide that folds into a native stable bold beta-hairpin in aqueous solution. *Nat. Struct. Biol.* **1994**, *1*, 584–590.
- (116) Muñoz, V.; Thompson, P. A.; Hofrichter, J.; Eaton, W. A. Folding Dynamics and Mechanism of  $\beta$ -Hairpin Formation. *Nature* **1997**, *390*, 196–199.
- (117) Schuler, B.; Eaton, W. A. Protein folding studied by single molecule. *Curr. Opin. Struct. Biol.* **2008**, *18*, 16–26.

- (118) Fesinmeyer, R. M.; Hudson, F. M.; Andersen, N. H. Enhanced hairpin stability through loop design: the case of the protein G B1 domain hairpin. *J. Am. Chem. Soc.* **2004**, *126*, 7238–7243.
- (119) Shen, Y.; Bax, A. SPARTA+: a modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network. *J. Biomol. NMR* **2010**, *48*, 13–22.
- (120) Tjandra, N.; Feller, S. E.; Pastor, R. W.; Bax, A. Rotational diffusion anisotropy of human ubiquitin from  $^{15}\text{N}$  NMR relaxation. *J. Am. Chem. Soc.* **1995**, *117*, 12562–12566.
- (121) Hodsdon, M. E.; Cistola, D. P. Backbone Mobility of Intestinal Fatty Acid-Binding Protein as Monitored by  $^{15}\text{N}$  NMR Relaxation and  $^1\text{H}$  Exchange. *Biochemistry* **1997**, *36*, 2278–2290.
- (122) Cornilescu, G.; Marquardt, J. L.; Ottiger, M.; Bax, A. Validation of Protein Structure from Anisotropic Carbonyl Chemical Shifts in a Dilute Liquid Crystalline Phase. *J. Am. Chem. Soc.* **1998**, *120*, 6836–6837.
- (123) Feng, W.; Tejero, R.; Zimmerman, D. E.; Inouye, M.; Montelione, G. T. Solution NMR Structure and Backbone Dynamics of the Major Cold-Shock Protein (CspA) from *Escherichia coli*: Evidence for Conformational Dynamics in the Single-Stranded RNA-Binding Site. *Biochemistry* **1998**, *37*, 10881–10896.
- (124) Vögeli, B.; Kazemi, S.; Güntert, P.; Riek, R. Spatial elucidation of motion in proteins by ensemble-based structure calculation using exact NOEs. *Nat. Struct. Mol. Biol.* **2012**, *19*, 1053–1057.
- (125) van der Vaart, A.; Merz, K. M., Jr. The Role of Polarization and Charge Transfer in the Solvation of Biomolecules. *J. Am. Chem. Soc.* **1999**, *121*, 9182–9190.
- (126) van der Vaart, A.; Bursulaya, B. D.; Brooks, C. L.; Merz, K. M. Are Many-Body Effects Important in Protein Folding? *J. Phys. Chem. B* **2000**, *104*, 9554–9563.
- (127) Macias, A. T.; MacKerell, A. D., Jr. CH/pi Interactions involving Aromatic Amino Acids: Refinement of the CHARMM Tryptophan Force Field. *J. Comput. Chem.* **2005**, *26*, 1452–1463.
- (128) Wang, S.; Orabi, E. A.; Baday, S.; Bernéche, S.; Lamoureux, G. Ammonium Transporters Achieve Charge Transfer by Fragmenting Their Substrate. *J. Am. Chem. Soc.* **2012**, *134*, 10419–10427.
- (129) Huang, L.; Roux, B. Automated Force Field Parameterization for Nonpolarizable and Polarizable Atomic Models Based on Ab Initio Target Data. *J. Chem. Theory Comput.* **2013**, *9*, 3543–3556.
- (130) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33–38.
- (131) Chothia, C.; Levitt, M.; Richardson, D. Helix to helix packing in proteins. *J. Mol. Biol.* **1981**, *145*, 215–250.
- (132) Lindhout, D. A.; Litowski, J. R.; Mercier, P.; Hodges, R. S.; Sykes, B. D. NMR solution structure of a highly stable de novo heterodimeric coiled-coil. *Biopolymers* **2004**, *75*, 367–375.
- (133) Jelsch, C.; Teeter, M. M.; Lamzin, V.; Pichon-Lesme, V.; Belssing, B.; Lecomte, C. Accurate Protein Crystallography at Ultra-High Resolution: Valence-Electron Distributions in Crambin. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 3171–3176.
- (134) Ulmer, T. S.; Ramirez, B. E.; Delaglio, F.; Bax, A. Evaluation of backbone proton positions and dynamics in a small protein by liquid crystal NMR spectroscopy. *J. Am. Chem. Soc.* **2003**, *125*, 9179–9191.
- (135) Schindelin, H.; Jiang, W.; Inouye, M.; Heinemann, U. Crystal structure of CspA, the major cold shock protein of *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91*, 5119–5123.
- (136) Vijay-Kumar, S.; Bugg, C. E.; Cook, W. J. Structure of Ubiquitin refined at 1.8 Angstroms Resolution. *J. Mol. Biol.* **1987**, *194*, 531–544.
- (137) Haglund, E.; Danielsson, J.; Kadhirvel, S.; Lindberg, M. O.; Logan, D. T.; Oliveberg, M. Trimming down a protein structure to its bare foldons: spatial organization of the cooperative unit. *J. Biol. Chem.* **2012**, *287*, 2731–2738.
- (138) Harata, K. X-ray structure of monoclinic turkey egg lysozyme at 1.3 Å resolution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1993**, *49*, 497–504.
- (139) Scapin, G.; Gordon, J. I.; Sacchettini, J. C. Refinement of the structure of recombinant rat intestinal fatty acid-binding apoprotein at 1.2-Å resolution. *J. Biol. Chem.* **1992**, *267*, 4253–4269.
- (140) Sandalova, T.; Schneider, G.; Kack, H.; Lindqvist, Y. Structure of Dethiobiotin Synthetase at 0.97 Å Resolution. *Acta Crystallogr., Sect. D* **1999**, *55*, 610–624.