HHS Public Access

Author manuscript

Methods Mol Biol. Author manuscript; available in PMC 2016 January 01.

Published in final edited form as:

Methods Mol Biol. 2015 ; 1215: 47–71. doi:10.1007/978-1-4939-1465-4_3.

Current Status of Protein Force Fields for Molecular Dynamics

Pedro E.M. Lopes[1], **Olgun Guvench**[2], and **Alexander D. MacKerell Jr.**[1]

[1]Department of Pharmaceutical Sciences University of Maryland School of Pharmacy 20 Penn Street HSFII Baltimore, Maryland 21201, USA

[2] Department of Pharmaceutical Sciences University of New England College of Pharmacy 716 Stevens Avenue Portland, Maine 04103, USA

Summary

The current status of classical force fields for proteins is reviewed. These include additive force fields as well as the latest developments in the Drude and AMOEBA polarizable force fields. Parametrization strategies developed specifically for the Drude force field are described and compared with the additive CHARMM36 force field. Results from molecular simulations of proteins and small peptides are summarized to illustrate the performance of the Drude and AMOEBA force fields.

Keywords

Force Field; Molecular Dynamics; Drude polarizable force field; CHARMM; AMOEBA; AMBER; GROMOS; OPLS; NAMD; electronic polarization

1. Introduction

Classical molecular dynamics (MD) simulations of proteins using empirical force fields have reached a mature state after 35 years of development and are now widely used as tools to investigate their structure and dynamics under a wide variety of conditions. These include studies of ligand binding, enzymatic-reaction mechanisms, protein folding and un-folding and protein-protein interactions.

Fundamental to such simulations is determination of the time evolution of the system's energy (protein for example) as a function of its atomic coordinates. An accurate description of the energy is thus required, since the lower energy states are expected to be populated. The gradient of the energy function, which is differentiable, is related to the forces acting on individual atoms. In chemistry the set of potential energy functions from which the forces are derived is commonly referred to as a force field (FF). As a result of many years of careful refinement, current additive protein energy functions are of sufficient quality that they may be used predicatively for studying protein dynamics and protein–protein interactions and in pharmacological applications (1). It is clear that the next major step in advancing protein force field accuracy requires a different representation of the molecular

Correspondence to: Alexander D. MacKerell, Jr..

energy surface. Specifically, the effects of charge polarization must be included, as fields induced by ions, solvent, other macromolecules, and the protein itself will affect electrostatic interactions (2-6).

Our goal here is to provide an update of the newest developments that have occurred in the field of FF-based MD simulations of proteins since our last review was published (7). Previously, we focused on the functional forms of additive FFs, strategies for parameter optimization, methodologies to perform MD simulations such as pressure and temperature control, and software packages available for MD simulations. Also briefly mentioned were efforts to extend additive FFs to other biomolecules. The FFs detailed were the Amber, CHARMM, GROMOS, and OPLS-AA additive protein FFs, with particular emphasis on CHARMM because of our continuing role in its development. While the present review begins with a brief update on the status of additive protein FFs, here we primarily focus on the latest developments in the inclusion of electronic polarizability into protein FFs. Emphasis is placed on the CHARMM Drude polarizable FF and the polarizable AMOEBA FF, and we direct interested readers to other recent reviews (5, 6, 8, 9). Also of interest may be new improvements in the Amber family of FFs (10), and, to our knowledge, no new reviews on the OPLS-AA or GROMOS protein FFs have appeared since our previous review in this series.

A general familiarity with molecular mechanics and dynamics and their applications to proteins is assumed. Simulation methods for proteins are well established, with many good textbooks and monographs covering the basics (11-17). The reader is also referred to chapter one of this volume.

2. Current Status of Additive Force Fields

Since the last review in this series (7), some notable developments have been made to additive FFs for proteins. Below, brief descriptions of the improvements introduced to two of the major additive FFs for proteins, CHARMM and Amber, are given.

2.1 CHARMM force field

The CHARMM additive all-atom FF has been in development since the early 80s (18) and has achieved a substantial degree of completeness with regard to coverage of chemical space. Apart from proteins (19, 20), it supports nucleic acids (21-23), lipids (24-26), and carbohydrates (27-30), allowing simulations on all commonly encountered motifs in biological systems. It has also been extended to cover the wide range of the chemical space required to study compounds common in medicinal chemistry through the CHARMM General FF (CGenFF) (31). The CHARMM additive FF for proteins recently underwent a significant update that culminated in the C36 version of the FF, as detailed below (20).

Long simulations with the additive C22/CMAP FF (19, 32, 33) had shown that certain fast-folding proteins would reach the native state, when started from a completely unfolded configuration (e.g. Villin headpiece subdomain) (34). However, significant deficiencies were also found. Examples of problems included misfolding encountered in long simulations of the pin WW domain and differences in the Villin folding mechanism from the

experimental results (34, 35). In the case of the WW domain, free energy calculations showed that the misfolded states had lower free energies than the folded state, confirming that the energy function could be further improved (36). A number of studies had suggested that such differences could be the result of small inaccuracies in the energy of the backbone, resulting in one structure being favored, and that this behavior could be corrected with minor adjustments to the backbone potential (37-40). Best *et al.* reported a revised set of CHARMM all-atom protein FF parameters (C36) that represents a significant improvement in the potential energy surface, while keeping the same functional form (20). The improvements that were introduced included (1) a new backbone CMAP potential, optimized against experimental data on small peptides and larger, folded proteins and (2) new side-chain dihedral parameters optimized using QM energies for dipeptides, conformational sampling in the model system $(Ala)_4$-X-$(Ala)_4$ (41) and NMR data from unfolded proteins. Other improvements relative to the previous C22/CMAP protein FF included Lennard-Jones (LJ) parameters for aliphatic hydrogens (42), internal parameters for the guanidinium ion (43), and improved parameters for tryptophan (44). Changes of the backbone and side-chains were done simultaneously, ensuring that in the new FF their contribution to protein structure and dynamics is balanced.

2.2 Amber force field

Amber FFs for proteins have been continually improved in recent years and a detailed discussion of the various changes is beyond the scope of this review. Significant revisions have been published, with particular emphasis on important dihedral angles. Simmerling and co-workers (45) introduced changes to the backbone potential in the original Amber ff99 FF by fitting to additional quantum-level data to produce the improved Amber ff99SB FF. Best and Hummer continued along the same line, modifying the backbone potential of the ff99SB and ff03 FFs to obtain a better balance between sampling of helix and coil conformations. The new FFs were named ff99SB* and ff03*, respectively (38). Modifications of the side-chain torsion potential for four amino acid types in ff99SB was introduced by Lindorff-Larsen *et al.* originating the ff99SB-ILDN FF (46). Further enhancements were produced by Li and Bruschweiler based on experimental NMR data, originating the ff99SB-ILDN-NMR FF (47). To our knowledge, the latest update in the Amber FFs was introduced recently by Neremberg and Head-Gordon, who included a perturbation to the φ backbone dihedral potential to shift the beta–PPII equilibrium. This resulted in improved sampling in water (TIP3P and TIP4P-Ew). Their updates were designated ff99SB-ILDN-Phi (48). In addition to proteins, the Amber FFs support most common biomolecules. The ff10 FF collection includes the most commonly used variants: the ff99SB protein parameters (45), the BSC0 DNA parameters (49), the Cheatham *et al.* ion parameters (50, 51), and updated RNA parameters (52, 53). Carbohydrates are supported through the Glycam FFs (54-56), and phospholipids are supported through the CHARMM FF and the recent Lipid11 FF (57).

3. Polarizable force fields for biomolecules. Current status

3.1 Drude Polarizable Force Field

Development of the Drude polarizable FF in CHARMM (58) started in 2001 and the capability to simulate the Drude model is now included in NAMD (59), ChemShell

QM/MM (60) and the OpenMM suite of utilities for GPUs (61). Development of the force field first involved implementation of the appropriate integrators to allow computationally efficient extended Langrangian MD simulations (62). This was followed by optimization of the first water model, in which a positive charge was assigned to the Drude particle (SWM4-DP) (63). The SWM4-DP model was re-optimized with a negative charged assigned to the Drude particles, consistent with their representation of the electronic degrees of freedom. The new model, called SWM4-NDP, is the standard polarizable water model of the Drude polarizable FF (64). It was calibrated to reproduce important properties of the neat liquid at room temperature and pressure such as enthalpy of vaporization, density, static dielectric constant and self-diffusion constant, free energy of hydration and shear viscosity. Concurrently with development of the water model, methodologies to determine electrostatic parameters for the Drude FF were advanced (65).

An early test of the feasibility of molecular dynamics simulations with the Drude polarizable FF was a successful simulation of a DNA octamer in a box of water with sodium counterions (66). Development of the Drude polarizable FF continued with parametrization of small molecules covering the functional groups commonly found in biomolecules. In 2005, the alkane FF was developed, followed by parametrization of alcohols and aromatic compounds in 2007 (67, 68). Harder *et al.* published the first generation of N-methyl acetamide (NMA) parameters in 2008 (69). Noteworthy is the proper treatment of the dielectric constant by the polarizable FF in all systems, a property considered essential for the accurate treatment of, for example, hydrophobic solvation in biomolecules. The Drude polarizable FF was extended to the nitrogen-containing heteroaromatic compounds in 2009 (5). FF parameters were refitted for ethers by Baker and MacKerell (70), with significant improvements in the reproduction of liquid phase dielectric constants, while maintaining the good agreement of the previous model with all other experimental and quantum mechanical target data (71). Sulfur containing model compounds were parametrized in 2010 (72). Other classes of molecules for which Drude empirical FF parameters had been developed are nucleic acid bases (73) and acyclic polyalcohols (74). Early simulations of dipalmitoylphosphatidylcholine (DPPC) bilayers and monolayers were reported (75), followed by completion of a refined model for DPPC (76).

Significant progress has been made in extending the Drude polarizable FF from small compounds representative of the building blocks encountered in biological polymers to the polymers themselves. The Drude empirical FF applicable to MD simulation studies of peptides and proteins, termed Drude-2013, is covered in Section 5, which includes a full account of the results. The optimization of the polypeptide backbone parameters is discussed in detail in Section 4.2 and the optimization of side-chain torsions is discussed in Section 4.3.

3.2 Amoeba force field

AMOEBA is another classical polarizable FF that has achieved the goal of producing a fully functional FF model for proteins (77). Development of the AMOEBA polarizable FF has been ongoing since 1995 (78) and is based on modeling the electrostatic energy using permanent and induced contributions. Permanent electrostatics originate in atomic

multipole-multipole interactions with moments up to the quadrupole located on each atom. The induced contribution is modeled iteratively by generating an induced dipole originated by permanent multipoles and other induced dipoles. Self-consistency is obtained using an iterative scheme, and the Thole model (79) is used to dampen electrostatic interactions at short range.

The AMOEBA FF was initially developed for water (80, 81). Testing included reproduction of a variety of experimental data and quantum calculations for small clusters, liquid water, and ice. Several liquid phase properties including bulk thermodynamic, transport, and structural measures were tested. These included density, heat of vaporization, self-diffusion coefficient, heat capacity, dielectric constant, and radial distribution functions. Overall, excellent agreement with reference values was obtained, and the model was demonstrated to be applicable to structural properties of two ice forms (81).

Treatment of ions in AMOEBA is described in reference (82). Absolute solvation free energies for potassium, sodium, and chloride ions in liquid water and formamide have been computed. Simulation results accurately reproduced vacuum QM results, experimental ion-cluster solvation enthalpies, and experimental solvation free energies for whole salts.

The AMOEBA FF has been extended to organic molecules, including alkanes, alcohols, amines, sulfides, aldehydes, carboxylic acids, amides, aromatics, and other small organic molecules (83). As a validation, the hydrogen bonding energies and structures of gas phase heterodimers with water were evaluated. Liquid self diffusion and static dielectric constants computed from MD simulations with AMOEBA are consistent with experimental values. The FF was further tested by computing the solvation free energy of 27 compounds not included in the parametrization process. It performed well across different environments and phases, yielding an RMS error of 0.69 kcal/mol. Analysis of the dependence of computed hydration free energies for seven small organic molecules with the QM level of theory used to derive atomic multipoles was presented recently (84). It was concluded that inclusion of diffuse functions in the QM calculation of the atomic multipoles is important. More comprehensive descriptions of the AMOEBA FF have been presented previously and the reader is referred to those publications for additional details (85-87).

4. Parametrization of polarizable force fields

4.1 Generic parametrization strategies for the Drude polarizable force field

The quality of FFs is heavily dependent on the quality of the underlying parameters. To obtain parameters of sufficient quality that are capable of producing accurate simulation results, procedures have been developed to target properties such as molecular geometries and vibrations, pure solvent properties, and free energies of solvation, among others during the parametrization. In this section we will describe parametrization of the polarizable Drude FF implemented in CHARMM. Reference to the well-establish protocol used to derive CHARMM additive FF parameters will be done whenever a parallel is useful. The general outline of the parametrization process has been described for the CHARMM additive FF in several publications (see references (1) and (19) for more details). Note that parameter optimization remains an iterative process in the polarizable FF and several rounds of

parametrization are typically performed until a satisfactory level of agreement with target data is obtained.

A common strategy in parameter optimization of biological macromolecules is that parameters are developed for small, representative model compounds and then transferred to the larger macromolecules. The advantages of this approach are: (1) smaller models are easier to treat using both MM and QM methods and (2) more experimental data are available for the smaller systems, including thermodynamic properties of condensed phases, such as heats of vaporization or sublimation and free energies of aqueous solvation. It is crucial to include such data in the parameter optimization process to get an accurate description of the non-bond portion of the FF. This strategy was also attempted in the development of Drude FF parameters for the protein backbone, but ultimately a more involved procedure was required as detailed in Section 4.2.

Parametrization of CHARMM FFs relies on obtaining appropriate intramolecular (bond, angle, dihedral, Urey-Bradley, and improper terms), van der Waals (vdW), and electrostatic parameters that adequately reproduce selected target data. Determination of the electrostatic parameters differs between the additive and the Drude polarizable FFs. In the Drude FFs, in addition to optimization of point charges, which is also required in the additive FF, polarizabilities and Thole factors must also be determined. In the additive CHARMM FF, optimization of point charges is based on a supramolecular approach where the charges are adjusted to reproduce QM HF/6-31G* interaction energies and geometries of the model compound with, typically, individual water molecules. Placement of water molecules at different orientations around the molecule enforces that local electronic polarization is accounted for implicitly, an important feature for accurate reproduction of condensed-phase properties. Additional data, often includes QM results on dimers and dipole moments of the models. It is well-known that in additive force fields, dipole moments must be overestimated to reproduce condensed phase properties (19, 88).

Other additive biomolecular FFs, most notably Amber, determine atomic partial charges based on reproduction of the QM Electrostatic Potential (ESP), evaluated on grids surrounding model compounds (89-91). These methods are convenient because charges can be developed quickly for any compound for which the QM ESP can be determined. An extension of ESP methods is inclusion of restraints during fitting, referred to as the restrained ESP (RESP) approach (92). This overcomes limitations on the determination of charges on buried atoms (93). It is important to note that partial charges from both supramolecular and ESP approaches are conformation-dependent, requiring care in the selection of appropriate conformations when performing the charge optimization.

Electrostatic parameters of model compounds in the Drude polarizable FF are obtained from restrained fitting to perturbed QM ESP maps on grid points located on concentric Connolly surfaces surrounding the molecule. Often fitting is supplemented with reproduction of the molecular dipole moment and diagonal elements of the polarization tensor (65, 94). The determination of the atomic polarizabilities and Thole factors (79) requires multiple perturbed ESPs typically calculated at the B3LYP/aug-cc-pVDZ level, with each giving the electronic response of the molecule to a point charge. Perturbing ions are placed mainly

along chemical bonds and lone pairs (LPs). This protocol was later extended to incorporate additional lone pair parameters and polarizability anisotropy, and has become the standard in developing electrostatic parameters for small molecules (95). LPs typically carry the charge of the atom (e.g., N, O, S in proteins) to which they are attached. The associated polarizability and Thole factor are both assigned to the parent atom. Anisotropic polarizability of hydrogen bond acceptors was found to be required to reproduce interactions with ions as a function of orientation. Initial values for the partial atomic charges are taken from the C22 additive all-atom FF, and those for the polarizabilities are based on adjusted Miller's atomic hybrid polarizability (ahp) values (96).

Although gas-phase properties (eg. dipole moments) are easily reproduced with full atomic polarizabilities, scaling of the polarizabilities has been shown to be necessary to reproduce condensed-phase properties (64). A scaling factor of approximately 0.7 was found appropriate for the SWM4-DP and SWM4-NDP water models while for other classes of molecules scaling factors range from 0.6 to 1.0, with 1.0 being full polarizability. For instance, scaling factors are 0.7 for primary and secondary alcohols (67), 0.85 for aromatics (68), N-containing heterocycles (94), nucleic acid bases (73) and ethers (97), and 1.0 for alkanes (42). Other scaling factors are 0.7 for thiols, 0.85 for dimethyl disulfide and 0.6 for ethylmethyl sulfide (72). A value of 0.724 was recently used in ion parameters (98). Final optimization of the electrostatic parameters consists of testing the model for reproduction of the pure solvent dielectric constants and adjusting the polarizability scaling if necessary.

Development of parameters to model vdW forces in the Drude FF, which are treated using the Lennard-Jones (LJ) 6-12 term, follows closely the protocol established for the additive FF and will only be briefly outlined here. Jorgensen and co-workers (99, 100) pioneered the use of condensed-phase simulations, usually pure liquids, as the basis for optimization of Lennard-Jones (LJ) parameters that account for both vdW attraction and inter-atomic repulsion. Typically, once electrostatic parameters are determined, the LJ parameters for a model compound can be adjusted to reproduce experimental pure solvent properties such as heat of vaporization, density, isothermal compressibility, heat capacity, heat of sublimation, lattice geometry, and free energy of aqueous solvation, as available. Although this is an effective method for the fine-tuning of the parameters, there are important issues. One is parameter correlation, such that LJ parameters for different atoms in a molecule and/or the magnitudes of ε_{ij} and $R_{\min}$ on the same atom, can compensate for individual unbalanced values, making it difficult to gauge whether they are balanced relative to one another (101). To overcome this problem, a method has been developed to determine the relative value of the LJ parameters based on high level QM data (102) with the absolute values being based on scans of ε_{ij} and $R_{\min}$ that reproduce experimental data (103, 104). This approach requires supramolecular interactions between rare gases and the model compound. Importantly, once satisfactory LJ parameters are obtained for atoms in a class of functional groups, they can often be directly transferred to other molecules carrying those functional groups without additional optimization.

Reproduction of experimental hydration free energies reflects how well the electrostatic and vdW parameters model interactions with bulk water. Recently, in the context of the polarizable Drude FF, it was shown that atom-pair-specific LJ parameters (termed “NBFix”

in the context of CHARMM) needed to be used in order to minimize discrepancies between calculated and experimental hydration free energies while simultaneously reproducing pure solvent heats of vaporization and molecular volumes (105).

Optimization of internal parameters is usually done relative to target data that include geometries, vibrational spectra and conformational energies. Geometries are typically optimized at the MP2/6-31G* level (or MP2/6-31+G* in the case of anions), and vibrational spectra are obtained at the MP2/6-31G* level. Frequencies are scaled using correction factors prescribed by Radom and co-workers (106), and a symbolic potential energy distribution (PED) analysis is performed as proposed by Pulay *et al*. (107) using the MOLVIB module in CHARMM. This approach has been shown to yield good agreement with experimental geometries for model compounds of complex systems such as proteins, nucleic acid bases, and sugars (22, 108, 109), while being computationally feasible.

4.2 Parametrization of the polypeptide backbone in the Drude force field

Parameterization of polypeptide backbone was initially assumed to follow the general rules in use for CHARMM FFs, namely that parameters would be transferable from smaller model compounds. The prototype of the protein backbone, for all residues except glycine and proline, was based on alanine polypeptides. The initial electrostatic model, identified as Drude-NMA, was derived from a combination of electrostatic parameters that included N-methyl acetamide (NMA) and ethane, and LJ parameters were also transferred from NMA and ethane. Several rounds of optimization were previously done on NMA: initial parameters were published by Harder *et al*. (69) and a final set by Lin *et al*. (110) In the latest model LJ parameters were selected to give acceptable intramolecular hydrogen bond distances in α-helix conformations of alanine polypeptides in addition to allowing reproduction of NMA experimental condensed phase properties (110). CMAP corrections for alanine dipeptide were also used to allow the (φ, ψ) Ramachandran map to reproduce a high-level QM (RIMP2/CBS//RIMP2/cc-pVDZ) surface, where the CBS (complete basis set) extrapolation was obtained from RIMP2/cc-pVTZ and RIMP2/cc-pVQZ single point energies following the prescription of Halkier *et al*. (111). The Drude-NMA model was tested by calculating gas phase molecular properties of alanine dipeptide and $(Ala)_5$ in different conformations, such as dipole moments, relative energies, and molecular polarizabilities, and through MD simulations of $(Ala)_5$ in solution (112). Testing the behavior of $(Ala)_5$ in solution has become common practice in the validation of protein force fields, being used in the development of the C36 additive FF (20) and the AMOEBA polarizable FF (77) (see Section 5.1) for details.

As alluded to above, direct transfer of Drude-NMA parameters to polypeptides did not yield acceptably accurate results (Table 1). Using transferred parameters, the agreement of the computed dipole moments, polarizabilities, and relative energies with target values was poor, in particular for the extended conformations. Tests of NMR *J*-coupling also indicated poor agreement with experiment due to a (φ, ψ) distribution that predominantly populates extended C5 conformations.

Included in Figure 1 are representative orientations and magnitudes of the induced dipoles and separations (pm) of the Drude particle and main atom in a dipeptide section of the

alanine dipeptide (values in parenthesis) and $(Ala)_5$. All calculations were done enforcing the C5 conformation for both systems. The separation between the Drude particle and the main atom is a direct measurement of the magnitude of the induced dipole. Using Drude-NMA electrostatic parameters (Figure 1A), displacement of the Drude particles in $(Ala)_5$ relative to the parent atom are similar on all carbonyl C (labeled C_{i-1}) atoms (~14 pm), and are substantially larger than the displacement for C_{i-1} in alanine dipeptide. C_{i-1} in alanine dipeptide is bound to a methyl group and the polarizing field is much weaker than in the longer polypeptide where C_{i-1} feels the electric field originating from the same amino acid's NH group. The case is similar for the N atoms, with N_{i+1} showing a much stronger induced local dipole in $(Ala)_5$ as compared to the alanine dipeptide. The induced dipole on C_α is smaller on $(Ala)_5$, enhancing the dipole interaction between N_i and C_i. This results in two effects. First, local dipoles associated with the peptide bonds interact with each other enhancing the local dipole moments associated with each peptide bond and, second, the larger dipole strengthens electrostatic interactions with water leading to overstabilization of the C5 conformation. Indeed, a comparison of the dipole moments of acetyl-$(Ala)_5$-N-methylamide for the NMA based model with QM data indicated the overall dipole moment of the C5 conformation to be significantly overestimated (Table 1). 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 defining balanced electrostatic interactions between the individual peptide bonds. Based on this analysis it was concluded that use of larger model compounds allowing communication between adjacent peptide bonds was required in the determination of electrostatic parameters, with the initial candidate being the alanine dipeptide.

Electrostatic parameters based on the alanine dipeptide were determined by averaging the components over five independent sets of parameters obtained from electrostatic potential (ESP) fitting corresponding to the αR, αL, C5, PPII and C7eq conformations. This model is referred to as Drude-ALA in the text below. For each conformation the electrostatic parameter optimization, which included the partial atomic charges, atomic polarizabilities, and atom-based Thole factors, was performed using the FITCHARGE module of CHARMM by fitting to the QM ESP maps as described above. The outcome is electrostatic parameters that better reproduce the change in the ESP associated with electrostatic interactions between the peptides bonds in the different relative orientations. The resulting Drude-ALA model yielded a smaller dipole moment for the C5 conformation for acetyl-$(Ala)_5$-N-methylamide (Table 1). Simulations of $(Ala)_5$ in aqueous solution were also performed and compared to Drude-NMA, and while the Drude-ALA model showed improved agreement with experiment, the agreement was still poor as compared to the additive C36 FF. It was found that the PPII region started to be populated, though the C5 conformation still dominated, indicating that the inclusion of electrostatic interactions between the peptide bonds during parameter optimization did improve the quality of the FF. However, those improvements were clearly insufficient, indicating that different target data were needed to obtain a more accurate electrostatic model for the polypeptide backbone.

The inability of Drude-ALA electrostatics to provide a reasonable description of properties of alanine polypeptides in gas-phase and solution prompted development of a third parametrization strategy. The rationale of the new methodology has its roots in the

fundamental physics of the Drude model. In the presence of an electric field the position of the Drude particles are optimized while the main atom remains fixed following the Born-Oppenheimer principle. This creates on each atom a local dipole that, although small, is able to interact with neighboring dipoles. The magnitude of each dipole can be controlled by two factors: (1) the atomic polarizability, and (2) the damping of 1-2 and 1-3 interactions through the individual Thole factors. Thus, for polymeric structures, control of the behavior of each dipole is extremely complex. Boundaries cannot be explicitly imposed locally for each atom or groups of atoms as in other polarizable methods since the overall properties are the result of many cross interactions spanning wide regions of the system. As a consequence, an effective methodology of parametrization needed to include enough information on the whole molecule, thus allowing for a balanced set of electrostatic parameters. Furthermore, it was necessary to include information not only in gas phase but also from interactions with water molecules, since water is the preferred medium where most of the MD simulations with Drude oscillators are anticipated to take place. The third optimization method of backbone electrostatic parameters used a Simulated Annealing (SA) protocol (113), yielding the final model, Drude-2013. The target data consisted of an array of QM observables determined for the alanine dipeptide and larger alanine polypeptides. Target data included the polarizability of the alanine dipeptide, relative energies of $(Ala)_5$, dipole moments of alanine dipeptide and $(Ala)_5$, and energetic and structural data for the interaction of the alanine dipeptide with individual water molecules along specific directions. Several conformations of the alanine models were used: αR, C5, and PPII for the relative energies of $(Ala)_5$; C5 and PPII for the interactions of the alanine dipeptide with water; and αR, C5, PPII, and C7eq conformations of the alanine dipeptide for molecular polarizabilities and dipole moments. In addition to the electrostatic parameters, during the SA internal parameters were allowed to vary within a limited range to keep the alanine dipeptide optimized geometries close to the targeted values. SA 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 criterion, resulting in Monte Carlo Simulated Annealing (MCSA) (114). The temperature was gradually reduced to near 0 K yielding a selected parameter set for testing in $(Ala)_5$ solution simulations. The error function was the weighted 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 (φ, ψ) RIMP2/CBS//MP2/6-311G(d,p) surface was generated at each iteration. In addition, empirical adjustments of the CMAP were added to the QM-based surface to improve agreement with conformational sampling of the peptide backbone in peptides and proteins, resulting in the final Drude-2013 model.

While both C36 and Drude-2013 (φ, ψ) surfaces have undergone some empirical adjustments, the underlying energy surfaces was based on quantum mechanics, and therefore the overall landscape of the surfaces is similar. Adjustments in the C36 CMAP, which was obtained at the LMP2/cc-pVQZ level, included local optimization of the helical and sheet regions to reproduce subtle features observed in crystallographic survey data (32) followed by subsequent shifting of the helical region to decrease the tendency for the C22/CMAP model to over-populate that conformation (20). For the Drude-2013 model, the overall sheet

region was lowered and the areas between the sheet and helical regions and from $\varphi = -90$ to -180 and $\psi = -60$ to 105° were raised.

4.3 Side chain χ1, χ2 dihedral parameter optimization in the Drude force field

Different side chains impact the conformational distribution of the polypeptide backbone, as observed in experimental studies (115-117). The peptide $(Ala)_4$-X- $(Ala)_4$ has been used before as a model system for χ1, χ2 parameter optimization (46), where X is the amino acid of interest and the backbone conformation is constrained to fully extended, C7eq, PPII, or αR conformations (41). Those studies indicated that $(Ala)_4$-X- $(Ala)_4$ with either the C7eq or PPII backbone conformation yields aqueous phase conformational properties that mimic those occurring in full proteins. Based on this analysis, χ1, χ2 parameter optimization was performed by initially targeting QM data for the respective side chain dipeptides, with the backbone in the β, αR, and αL conformations. These parameters were then used in Hamiltonian Replica Exchange MD (H-REMD) simulations (118) of $(Ala)_4$-X- $(Ala)_4$ in solution, with χ1, χ2 sampling compared with PDB survey data. Overlap coefficients (OC) (41) for χ1 and χ2 distributions from $(Ala)_4$-X- $(Ala)_4$ in the C7eq conformation and those from a survey of the PDB (119) were computed, with an OC of 1 indicating exact agreement and an OC of 0 indicating no agreement. The extent of overlap for some of the amino acids based on optimization only targeting the QM data was found to be quite good. For example, values of 0.87, 0.88 and 0.87 were obtained for χ1 for Cys, Leu, and Val, respectively, while the OC was 0.92 for χ2 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 χ1 and χ2 populations of the gauche+, gauche−, and trans rotamers and manually adjusting the corresponding dihedral parameters to improve the level of agreement. After the optimization, significant agreement with the PDB target data was obtained 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 were also found including Asn χ2, Asp χ1, Gln χ2, and Glu χ1. The final parameters were used for the reported polypeptide and protein simulations. In reference (120) we present detailed descriptions of the optimization protocol and final results.

4.4 The AMOEBA force field and parametrization of proteins

Detailed methodology for deriving electrostatic parameters for AMOEBA to allow incorporation of novel molecules has been published (83), and therefore what follows is a brief overview. Determination of permanent atomic multipoles for glycine, alanine, and proline residues was done based on capped acetyl-X-N-methylamide dipeptides with X = Gly, Ala, and Pro. The first step is definition of intramolecular direct polarization groups, which is important because atoms belonging to one group can only polarize atoms outside that group. The group definitions for alanine dipeptide are show in Figure 2 of reference (77). For side chains, groups are also selected. The optimization proceeds with assignment of the initial multipole parameters from Distributed Multipole Analysis (DMA) at the MP2/6-311G** level. Initial parameters are then iteratively optimized against the MP2/aug-cc-pVTZ electrostatic potential computed on a set of grid points around the dipeptide compounds. Converged Permanent Atomic Multipoles (PAMs) were determined simultaneously for five local minima: αL, α', C5, C7a, and C7e conformers

5. Application of polarizable force fields to protein simulations

The year 2013 marked important milestones in the development of polarizable FFs. After years of development, polarizable FFs for peptides and proteins suitable for MD simulations based on classical Drude oscillators (Drude-2013) and the AMOEBA model (AMOEBA-2013) were published. Here, we summarize results of MD simulations with the two FFs.

5.1 Peptide simulations with C36 additive, AMOEBA-2013, and Drude-2013 force fields

With advances in computing capacity, it has become common to use simulations of oligopeptides in solution to calibrate FF torsional parameters (20, 45, 112, 121-124), since results can be directly compared to experimental nuclear magnetic resonance (NMR) data for corresponding peptides. Conformational distributions in an NMR experiment are reflected in NMR-derived spin–spin coupling (*J*-coupling) constants. Using Karplus relations, *J*-coupling values can be computed from peptide conformations from MD simulations, and the ability to achieve ever-increasing timescales via MD allows for the computational generation of conformational ensembles of a size that can be meaningfully compared with experiment (121, 125).

As an example, simulations of small polypeptides of $(Ala)_3$, $(Ala)_5$, $(Ala)_7$, $(Val)_3$ and $(Gly)_3$ were used by Best *et al* to validate the improved CHARMM36 additive FF (C36) (20). Using this approach, C36 introduced small but significant changes relative to its predecessor, C22/CMAP. In alanine- and valine-based peptides, minima occur at PPII, with C5 and αR being only slightly higher in energy. The additional minima at αL and C7ax are approximately 2–3 kcal/mol higher than the PPII conformation. And while there is only a small difference between $(Ala)_3$, $(Ala)_5$, and $(Ala)_7$, sampling for $(Val)_3$ was significantly different because of the presence of the bulky hydrophobic side chain. Compared with other FFs, AMBER ff99SB9 and ff99SB* are closest to C36, while OPLS/AA (126) is qualitatively different with a minimum at C7eq and Gromos 53a6 FF (127) has two minima near αR and a low-energy transition region between αR and C7ax.

C36 (φ, ψ) sampling has also been compared with experimental NMR *J*-coupling. Agreement was very good for the alanine-based peptides and for $(Gly)_3$, and reasonable for $(Val)_3$. The new C36 FF significantly improves over the previous C22/CMAP FF, with improvement coming from decreased sampling of αR conformations and increased sampling of PPII, which is reflected in the *J*-couplings. In reference (20) C36 was also compared with other FFs outside the CHARMM family (AMBER ff99SB (45), OPLS/AA (126), Gromos 53a6 (127)), showing significantly lower χ^2 values. No direct comparison of C36 with the latest improved Amber FFs has been published, although it is anticipated that C36 will compare very favorably to experimental data based on published results.

Sampling for the unblocked, protonated $(Ala)_5$ peptide has been tested using the AMOEBA-2013 FF (77). Sampling is similar to C36, with a distinct global minimum located around the PII conformation and two other basins approximately 0.5 kcal/mol higher in free energy in the β-sheet and α-helix regions. Barriers separating the global basin from the two local minima are 1–2 kcal/mol. (φ, ψ) sampling was compared with experimental *J*-

coupling constants, and values from MD simulations are in excellent agreement with those probed by experiment with a χ^2 value of 1.0

Simulations of $(Ala)_5$ polypeptides were not used to validate the newly developed Drude polarizable FF in CHARMM but rather were explicitly part of the optimization process, particularly for the fine tuning of the CMAP potential so as to yield acceptable sampling patterns in the tested protein systems. In addition, the GB1 (41-56) hairpin (128, 129) and a dimeric coiled coil (1UOI) (130) were also used as target data for optimization of the Drude-2013 model. Due to the use of multiple small peptides, as well as larger proteins, as target data, sampling of $(Ala)_5$ had to be slightly compromised, yielding χ^2 values larger than 1.0.

Explicit solvent simulations of the GB1 hairpin of 100 ns yielded RMS differences with the Drude model more similar to the crystal structure of the full GB1 protein as compared to the C36 where the RMS difference fluctuated between 2.5 and 3 Å, indicating drift away from the crystal structure. With the dimeric coiled coil (1UOI) (130) RMS analysis showed the overall structure of the coiled coil to deviate more from the crystal structure with the Drude model as compared to C36 (see reference (120) for details). The individual helices in the dimer move relative to each other, while the conformations of the individual helices are well preserved, suggesting the ability of the Drude-2013 model to properly treat the helical secondary structure of the individual monomers. (φ, ψ) probability distributions from the simulations supported this conclusion. Thus, the Drude model satisfactorily reproduces the conformational properties of small peptides on the 100 ns time scale, though longer simulations will be required to more rigorously challenge the model.

5.2 Full proteins

Further validation of the Drude-2013 force field involved explicit solvent MD simulations on 10 proteins: 1EJG (crambin), 1P7E (protein GB1 domain), 1MJC (cold-shock protein A), 1UBQ (ubiquitin), 3ZZP (circular permutant of ribosomal protein), 4IEJ (DNA methyl transferase associated protein), 135L (lysozyme), 1IFC (fatty acid binding protein), 3VQF (PDZ domain from tight junction regulatory protein), and 1BYI (dethiobiotin synthase). The proteins are relatively small, typically less than 100 residues, and cover a range of secondary structures.

The stability of each protein was characterized by the value of its backbone RMS deviation (RMSD) relative to the crystal structure. Results, summarized in Table 7 and Figure S2 of the Supplementary Information of reference (120) showed the RMS differences are typically larger with the Drude model versus C36 additive force field as are the RMS fluctuations. The Drude-2013 model shows additional flexibility compared to the additive model with only one exception, namely ubiquitin (1UBQ). While the Drude model generally appears to have more flexibility than the additive C36 model, NMR analysis indicated that for selected residues with high mobility in C36, the Drude model gave improved agreement with experiment, as shown in Figure 7 of Lopes et al. (120).

Results with AMOEBA-2013 protein FF (77) have been reported for three of the proteins studied with the Drude-2013 force field. These include crambin (1EJG), ubiquitin (1UBQ)

and lysozyme. AMOEBA MD simulations were performed for 30 ns yielding backbone RMSDs in the vicinity of 1, 2, and 2 Å for the three proteins, respectively. At 30 ns of the Drude simulations the corresponding values were 1.1/1.1, 1.9 and 1.9/2.2 Å, were two values are from duplicate simulations.

Although the Drude model showed additional flexibility over the additive C36 model, the overall structures of the proteins is well maintained. Snapshots taken at 100 ns for lysozyme (135L) and dethiobiotin synthase (1BYI) superimposed on the corresponding crystal structures are shown in Figure 2, showing the overall maintenance of the structures. Consistent with this was the (φ, ψ) sampling with the Drude model over all the simulated proteins being similar to a survey of the PDB as well as sampling occurring with C36 (Figure 6 of Lopes et al). In addition, the N-H...O=C distance distributions in secondary structures were reasonably reproduced by the Drude model, though there is a tendency towards the distances being slightly longer than distributions from PDB crystal structures (Figure 5 of Lopes et al.).

Additional analysis involved dipole moments of selected moieties during the MD simulations. These included the peptide bonds in the GB1 hairpin and ubiquitin and tryptophan residues in lysozyme (Figure 8 of Lopes et al). In all cases the Drude dipole moments are systematically larger than with the additive model. This indicates that, while partial atomic charges in the additive model are adjusted to overestimate molecular dipole moments, the extent of overestimation is not enough for the protein environment. In addition, the dipole moments of the peptide bonds with the Drude model in sheets are systematically larger than in helices. Finally, significant variations in the dipole moments were observed in the Drude simulations (eg. > 1.5 D for a Trp in lysozyme). Thus, even though the additive models were optimized to yield enhanced dipole moments appropriate for the condensed phase, it does not appear that the overestimation was sufficient based on these initial polarizable calculations. That, as well as the large variation in the dipoles occurring in the Drude model, suggest that the underlying physical forces dictating the overall properties of the peptides and proteins is significantly different in the Drude versus the additive model. Indeed, the additional flexibility in the Drude model may be due to the inclusion of electronic polarization in the model allowing for the variability of the local molecular dipoles.

6. Summary

The field of empirical FF based simulations of proteins continues to develop. Since the last publication of a similar review great progress has been made including the publication of two polarizable force fields for proteins as well as improvements in the AMBER and CHARMM additive protein force fields. Work on other classes of biopolymers has also made significant progress allowing for simulations of heterogeneous systems. As other researchers start using the recently published force fields, in particular the polarizable force fields, limitations will certainly be found and corrections and improvements are expected.

As was emphasized in this review, development of electrostatic parameters in the Drude force field is very complex. It is expected that new optimization algorithms together with

more sophisticated target data will lead to significant progress. Polarizable models for other classes of biomolecules based on the Drude oscillator will be published soon for DNA and carbohydrates as well as a wider range of lipids.

While polarizable MD simulations will make a significant contribution to our understanding of protein structure and function it should be emphasized that these models are more sensitive to initial conditions than with an additive FF, and can have polarization catastrophes that will cause simulations to fail. To overcome this it is suggested that systems initially be set up and equilibrated with an additive FF and then converted to the polarizable model. To facilitate this procedure the CHARMM-GUI (131) has been extended to include a new utility, the “Drude Prepper.” The Drude Prepper reads equilibrated CHARMM PSF and coordinate files and converts them to Drude format files. This includes the production of inputs for MD simulations using CHARMM or NAMD. This utility will greatly facilitate the application of the Drude model to a range of proteins as well as other systems.

Concerning computational efficiency, the Drude model typically requires the use of a 1 fs integration time step during MD simulations. In addition, there is an approximately 2-fold overhead associated with the calculation of the polarization contribution to the electrostatics. Thus, the model is approximately 4-fold slower than corresponding additive simulations performed with a 2 fs integration time step. However, the NAMD implementation is highly parallelizable (59), which will facilitate simulations of large systems using the Drude model.

Acknowledgement

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.

Bibliography

1. MacKerell AD. Empirical force fields for biological macromolecules: Overview and issues. Journal of Computational Chemistry. 2004; 25(13):1584–1604. [PubMed: 15264253]
2. Stone AJ. Intermolecular potentials. Science. 2008; 321(5890):787–789. [PubMed: 18687950]
3. Freddolino PL, Harrison CB, Liu YX, Schulten K. Challenges in protein-folding simulations. Nature Physics. 2010; 6(10):751–758. [PubMed: 21297873]
4. Warshel A, Kato M, Pisliakov AV. Polarizable force fields: History, test cases, and prospects. Journal of Chemical Theory and Computation. 2007; 3(6):2034–2045.
5. Lopes PEM, Roux B, MacKerell AD. Molecular modeling and dynamics studies with explicit inclusion of electronic polarizability: theory and applications. Theoretical Chemistry Accounts. 2009; 124(1-2):11–28. [PubMed: 20577578]
6. Zhu X, Lopes PEM, MacKerell AD. Recent developments and applications of the CHARMM force fields. Wiley Interdisciplinary Reviews-Computational Molecular Science. 2012; 2(1):167–185. [PubMed: 23066428]
7. Guvench, O.; MacKerell, AD. Comparison of Protein Force Fields for Molecular Dynamics Simulations.. In: Kukol, A., editor. Molecular Modeling of Proteins. Humana Press; 2008. p. 63-88.
8. Lopes, PEM.; Harder, E.; Roux, B.; MacKerell, AD. Formalisms for the Explicit Inclusion of Electronic Polarizability in Molecular Modeling and Dynamics Studies.. In: York, DM.; Lee, T-S., editors. Multi-scale Quantum Models for Biocatalysis. Springer; Netherlands: 2009. p. 219-257.
9. Salomon-Ferrer R, Case DA, Walker RC. An overview of the Amber biomolecular simulation package. Wiley Interdisciplinary Reviews: Computational Molecular Science. 3(2):198–210.

10. Beauchamp K, Lin Y-S, Das R, Pande V. Are Protein Force Fields Getting Better? A Systematic Benchmark on 524 Diverse NMR Measurements. Journal of Chemical Theory and Computation. 2012; 8(4):1409–1414. [PubMed: 22754404]
11. Burkert, U.; Allinger, N. Molecular mechanics. American Chemical Society; 1982.
12. McCammon, JAHSC. Dynamics of proteins and nucleic acids. Cambridge University Press; Cambridge [Cambridgeshire]; New York: 1987.
13. Leach, AR. Molecular modelling : principles and applications. Prentice Hall; Harlow, England; New York: 2001.
14. Becker, OM. Computational biochemistry and biophysics. M. Dekker; New York: 2001.
15. Rapaport, DC. The art of molecular dynamics simulation. Cambridge University Press; Cambridge, UK; New York, NY: 2004.
16. Schlick, T. Molecular modeling and simulation : an interdisciplinary guide. Springer; New York: 2002.
17. Satoh, A. Introduction to practice of molecular simulation molecular dynamics, Monte Carlo, Brownian dynamics, Lattice Boltzmann, dissipative particle dynamics. Available from: http://site.ebrary.com/id/10440534
18. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. Journal of Computational Chemistry. 1983; 4(2):187–217.
19. MacKerell AD, Bashford D, Bellott, Dunbrack RL, Evanseck JD, Field MJ, et al. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. The Journal of Physical Chemistry B. 1998; 102(18):3586–3616. [PubMed: 24889800]
20. Best RB, Zhu X, Shim J, Lopes PEM, Mittal J, Feig M, et al. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone φ, ψ and Side-Chain χ_1 and χ_2 Dihedral Angles. Journal of Chemical Theory and Computation. 2012; 8(9):3257–3273. [PubMed: 23341755]
21. MacKerell AD, Wiorkiewicz-Kuczera J, Karplus M. An all-atom empirical energy function for the simulation of nucleic acids. Journal of the American Chemical Society. 1995; 117(48):11946–11975.
22. Foloppe N, MacKerell AD. All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. Journal of Computational Chemistry. 2000; 21(2):86–104.
23. MacKerell AD, Banavali NK. All-atom empirical force field for nucleic acids: II. Application to molecular dynamics simulations of DNA and RNA in solution. Journal of Computational Chemistry. 2000; 21(2):105–120.
24. Feller SE, MacKerell AD. An Improved Empirical Potential Energy Function for Molecular Simulations of Phospholipids. The Journal of Physical Chemistry B. 2000; 104(31):7510–7515.
25. Feller SE, Gawrisch K, MacKerell AD. Polyunsaturated Fatty Acids in Lipid Bilayers: Intrinsic and Environmental Contributions to Their Unique Physical Properties. Journal of the American Chemical Society. 2001; 124(2):318–326. [PubMed: 11782184]
26. Klauda JB, Venable RM, Freites JA, O'Connor JW, Tobias DJ, Mondragon-Ramirez C, et al. Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. The Journal of Physical Chemistry B. 2010; 114(23):7830–7843. [PubMed: 20496934]
27. Kuttel M, Brady JW, Naidoo KJ. Carbohydrate solution simulations: Producing a force field with experimentally consistent primary alcohol rotational frequencies and populations. Journal of Computational Chemistry. 2002; 23(13):1236–1243. [PubMed: 12210149]
28. Guvench O, Greene SN, Kamath G, Brady JW, Venable RM, Pastor RW, et al. Additive empirical force field for hexopyranose monosaccharides. Journal of Computational Chemistry. 2008; 29(15):2543–2564. [PubMed: 18470966]
29. Hatcher ER, Guvench O, MacKerell AD. CHARMM Additive All-Atom Force Field for Acyclic Polyalcohols, Acyclic Carbohydrates, and Inositol. Journal of Chemical Theory and Computation. 2009; 5(5):1315–1327. [PubMed: 20160980]

30. Guvench O, Hatcher E, Venable RM, Pastor RW, MacKerell AD. CHARMM Additive All-Atom Force Field for Glycosidic Linkages between Hexopyranoses. Journal of Chemical Theory and Computation. 2009; 5(9):2353–2370. [PubMed: 20161005]
31. Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, et al. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. Journal of Computational Chemistry. 2010; 31(4):671–690. [PubMed: 19575467]
32. MacKerell AD, Feig M, Brooks CL. 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. Journal of Computational Chemistry. 2004; 25(11):1400–1415. [PubMed: 15185334]
33. MacKerell AD, Feig M, Brooks CL. Improved treatment of the protein backbone in empirical force fields. Journal of the American Chemical Society. 2004; 126(3):698–699. [PubMed: 14733527]
34. Freddolino PL, Schulten K. Common Structural Transitions in Explicit-Solvent Simulations of Villin Headpiece Folding. Biophysical Journal. 2009; 97(8):2338–2347. [PubMed: 19843466]
35. Freddolino PL, Liu F, Gruebele M, Schulten K. Ten-microsecond molecular dynamics simulation of a fast-folding WW domain. Biophysical Journal. 2008; 94(10):L75–L77. [PubMed: 18339748]
36. Freddolino PL, Park S, Roux B, Schulten K. Force Field Bias in Protein Folding Simulations. Biophysical Journal. 2009; 96(9):3772–3780. [PubMed: 19413983]
37. Best R, Buchete N-V, Hummer G. Are current molecular dynamics force fields too helical? Biophysical journal. 2008; 95(1):L07–L09. [PubMed: 18456823]
38. Best RB, Hummer G. Optimized Molecular Dynamics Force Fields Applied to the Helix-Coil Transition of Polypeptides. The Journal of Physical Chemistry B. 2009; 113(26):9004–9015. [PubMed: 19514729]
39. Best RB, Mittal J. Balance between α and β Structures in Ab Initio Protein Folding. Journal of Physical Chemistry B. 2010; 114(26):8790–8798.
40. Mittal J, Best RB. Tackling Force-Field Bias in Protein Folding Simulations: Folding of Villin HP35 and Pin WW Domains in Explicit Water. Biophysical Journal. 2010; 99(3):L26–L28. [PubMed: 20682244]
41. Shim J, Zhu X, Best RB, MacKerell AD. Ala_4-X-Ala_4 as a model system for the optimization of the χ_1 and χ_2 amino acid side-chain dihedral empirical force field parameters. Journal of Computational Chemistry. 2013; 34(7):593–603. [PubMed: 23197420]
42. Vorobyov IV, Anisimov VM, MacKerell AD. Polarizable empirical force field for alkanes based on the classical drude oscillator model. Journal of Physical Chemistry B. 2005; 109(40):18988–18999.
43. Mason PE, Neilson GW, Enderby JE, Saboungi ML, Dempsey CE, MacKerell AD, et al. The structure of aqueous guanidinium chloride solutions. Journal of the American Chemical Society. 2004; 126(37):11462–11470. [PubMed: 15366892]
44. Macias AT, MacKerell AD. CH/pi interactions involving aromatic amino acids: Refinement of the CHARMM tryptophan force field. Journal of Computational Chemistry. 2005; 26(14):1452–1463. [PubMed: 16088926]
45. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins: Structure, Function, and Bioinformatics. 2006; 65(3):712–725.
46. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis J, Dror R, et al. Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins. 2010; 78(8):1950–1958. [PubMed: 20408171]
47. Li D-W, Bruschweiler R. NMR-Based Protein Potentials. Angewandte Chemie. 2011; 122(38): 6930–6932.
48. Nerenberg P, Head-Gordon T. Optimizing Protein−Solvent Force Fields to Reproduce Intrinsic Conformational Preferences of Model Peptides. Journal of Chemical Theory and Computation. 2011; 7(4):1220–1230.

49. Perez A, Marchan I, Svozil D, Sponer J, Cheatham TE, Laughton CA, et al. Refinenement of the AMBER force field for nucleic acids: Improving the description of alpha/gamma conformers. Biophysical Journal. 2007; 92(11):3817–3829. [PubMed: 17351000]
50. Joung IS, Cheatham TE. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. Journal of Physical Chemistry B. 2008; 112(30): 9020–9041.
51. Joung IS, Cheatham TE. Molecular Dynamics Simulations of the Dynamic and Energetic Properties of Alkali and Halide Ions Using Water-Model-Specific Ion Parameters. Journal of Physical Chemistry B. 2009; 113(40):13279–13290.
52. Banas P, Hollas D, Zgarbova M, Jurecka P, Orozco M, Cheatham TE III, et al. Performance of Molecular Mechanics Force Fields for RNA Simulations: Stability of UUCG and GNRA Hairpins. Journal of Chemical Theory and Computation. 2010; 6(12):3836–3849.
53. Zgarbova M, Otyepka M, Sponer J, Mladek A, Banas P, Cheatham TE III, et al. Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. Journal of Chemical Theory and Computation. 2011; 7(9):2886–2902. [PubMed: 21921995]
54. Kirschner KN, Woods RJ. Solvent interactions determine carbohydrate conformation. Proceedings of the National Academy of Sciences of the United States of America. 2001; 98(19):10541–10545. [PubMed: 11526221]
55. Woods RJ, Dwek RA, Edge CJ, Fraser-Reid B. Molecular Mechanical and Molecular Dynamic Simulations of Glycoproteins and Oligosaccharides. 1. GLYCAM_93 Parameter Development. Journal of Physical Chemistry. 1995; 99(11):3832–3846.
56. Kirschner KN, Yongye AB, Tschampel SM, González-Outeiriño J, Daniels CR, Foley BL, et al. GLYCAM06: A generalizable biomolecular force field. Carbohydrates. Journal of Computational Chemistry. 2008; 29(4):622–655. [PubMed: 17849372]
57. Skjevik, Ã.g.A.; Madej, BD.; Walker, RC.; Teigen, K. LIPID11: A Modular Framework for Lipid Simulations Using Amber. The Journal of Physical Chemistry B. 2012; 116(36):11124–11136. [PubMed: 22916730]
58. Brooks BR, Brooks CL 3rd, MacKerell AD Jr. Nilsson L, Petrella RJ, Roux B, et al. CHARMM: the biomolecular simulation program. J. Comput. Chem. 2009; 30(10):1545–614. [PubMed: 19444816]
59. Jiang W, Hardy DJ, Phillips JC, Mackerell AD 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(2):87–92. [PubMed: 21572567]
60. Boulanger E, Thiel W. Solvent Boundary Potentials for Hybrid QM/MM Computations Using Classical Drude Oscillators: A Fully Polarizable Model. J. Chem.Theo. Comp. 2012; 8:4527–4538.
61. Eastman P, Friedrichs MS, Chodera JD, Radmer RJ, Bruns CM, Ku JP, et al. OpenMM 4: A Reusable, Extensible, Hardware Independent Library for High Performance Molecular Simulation. J. Chem. Theory Comp. 2012; 8:461–469.
62. Lamoureux G, Roux B. Modelling Induced Polarizability with Drude Oscillators: Theory and Molecular Dynamics Simulation Algorithm. J. Chem. Phys. 2003; 119:5185–5197.
63. Lamoureux G, MacKerell AD, Roux B. A simple polarizable model of water based on classical Drude oscillators. Journal of Chemical Physics. 2003; 119(10):5185–5197.
64. Lamoureux G, Harder E, Vorobyov IV, Roux B, MacKerell AD. A polarizable model of water for molecular dynamics simulations of biomolecules. Chemical Physics Letters. 2006; 418(1-3):245–249.
65. Anisimov VM, Lamoureux G, Vorobyov IV, Huang N, Roux B, MacKerell AD. Determination of electrostatic parameters for a polarizable force field based on the classical Drude oscillator. Journal of Chemical Theory and Computation. 2005; 1(1):153–168.
66. Anisimov VM, Vorobyov IV, Lamoureux G, Noskov S, Roux B, MacKerell AD. CHARMM all-atom polarizable force field parameter development for nucleic acids. Biophysical Journal. 2004; 86(1):415A–415A.

67. Anisimov VM, Vorobyov IV, Roux B, MacKerell AD. Polarizable empirical force field for the primary and secondary alcohol series based on the classical drude model. Journal of Chemical Theory and Computation. 2007; 3(6):1927–1946. [PubMed: 18802495]

68. Lopes PEM, Lamoureux G, Roux B, MacKerell AD. Polarizable empirical force field for aromatic compounds based on the classical drude oscillator. Journal of Physical Chemistry B. 2007; 111(11):2873–2885.

69. Harder E, Anisimov VM, Whitfield TW, MacKerell AD, Roux B. Understanding the dielectric properties of liquid amides from a polarizable force field. Journal of Physical Chemistry B. 2008; 112(11):3509–3521.

70. Baker CM, MacKerell AD. Polarizability rescaling and atom-based Thole scaling in the CHARMM Drude polarizable force field for ethers. Journal of Molecular Modeling. 2010; 16(3): 567–576. [PubMed: 19705172]

71. Vorobyov I, Anisimov VM, Greene S, Venable RM, Moser A, Pastor RW, et al. Additive and classical drude polarizable force fields for linear and cyclic ethers. Journal of Chemical Theory and Computation. 2007; 3(3):1120–1133.

72. Zhu X, MacKerell AD. Polarizable Empirical Force Field for Sulfur-Containing Compounds Based on the Classical Drude Oscillator Model. Journal of Computational Chemistry. 2010; 31(12): 2330–2341. [PubMed: 20575015]

73. Baker CM, Anisimov VM, MacKerell AD. Development of CHARMM Polarizable Force Field for Nucleic Acid Bases Based on the Classical Drude Oscillator Model. Journal of Physical Chemistry B. 2011; 115(3):580–596.

74. He X, Lopes PEM, MacKerell AD. Polarizable Empirical Force Field for Acyclic Polyalcohols Based on the Classical Drude Oscillator. Biopolymers. 2013; 99(10):724–738. [PubMed: 23703219]

75. Harder E, MacKerell AD, Roux B. Many-Body Polarization Effects and the Membrane Dipole Potential. Journal of the American Chemical Society. 2009; 131(8):2760–2761. [PubMed: 19199514]

76. Chowdhary J, Harder E, Lopes PEM, Huang L, MacKerell AD, Roux B. A Polarizable Force Field of Dipalmitoylphosphatidylcholine Based on the Classical Drude Model for Molecular Dynamics Simulations of Lipids. Journal of Physical Chemistry B. 2013; 117(31):9142–9160.

77. Shi Y, Xia Z, Zhang JH, Best RB, Wu C, Ponder JW, et al. Polarizable Atomic Multipole-Based AMOEBA Force Field for Proteins. Journal of Chemical Theory and Computation. 2013; 9(9): 4046–4063. [PubMed: 24163642]

78. Dudek MJ, Ponder JW. Accurate modeling of the intramolecular electrostatic energy of proteins. Journal of Computational Chemistry. 1995; 16(7):791–816.

79. Thole B. Molecular polarizabilities calculated with a modified dipole interaction. Chemical Physics. 1981; 59(3):341–350.

80. Ren PY, Ponder JW. Polarizable atomic multipole water model for molecular mechanics simulation. Journal of Physical Chemistry B. 2003; 107(24):5933–5947.

81. Ren PY, Ponder JW. Temperature and pressure dependence of the AMOEBA water model. Journal of Physical Chemistry B. 2004; 108(35):13427–13437.

82. Grossfield A, Ren PY, Ponder JW. Ion solvation thermodynamics from simulation with a polarizable force field. Journal of the American Chemical Society. 2003; 125(50):15671–15682. [PubMed: 14664617]

83. Ren P, Wu C, Ponder JW. Polarizable Atomic Multipole-Based Molecular Mechanics for Organic Molecules. Journal of Chemical Theory and Computation. 2011; 7(10):3143–3161. [PubMed: 22022236]

84. Shi Y, Wu C, Ponder JW, Ren P. Multipole Electrostatics in Hydration Free Energy Calculations. Journal of Computational Chemistry. 2011; 32(5):967–977. [PubMed: 20925089]

85. Ponder, JW.; Case, DA. Protein Simulations. Academic Press; San Diego: 2003. Force fields for protein simulations.; p. 27-85.

86. Ponder JW, Wu C, Ren P, Pande VS, Chodera JD, Schnieders MJ, et al. Current Status of the AMOEBA Polarizable Force Field. The Journal of Physical Chemistry B. 2010; 114(8):2549–2564. [PubMed: 20136072]

87. Ren PY, Ponder JW. Consistent treatment of inter- and intramolecular polarization in molecular mechanics calculations. Journal of Computational Chemistry. 2002; 23(16):1497–1506. [PubMed: 12395419]

88. Jorgensen WL, Tirado-Rives J. The OPLS potential function for proteins. Energy minimizations for crystals of cyclic peptides and crambin. Journal of the American Chemical Society. 1988; 110:1657–1666.

89. Singh UC, Kollman PA. An approach to computing electrostatic charges for molecules. Journal of Computational Chemistry. 1984; 5(2):129–145.

90. Chirlian LE, Francl MM. Atomic charges derived from electrostatic potentials: A detailed study. Journal of Computational Chemistry. 1987; 8(6):894–905.

91. Merz KM. Analysis of a large data base of electrostatic potential derived atomic charges. Journal of Computational Chemistry. 1992; 13(6):749–767.

92. Bayly CI, Cieplak P, Cornell WD, Kollman PA. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. Journal of Physical Chemistry. 1993; 97(40):10269–10280.

93. Francl M, Carey C, Chirlian L, Gange D. Charges fit to electrostatic potentials. II. Can atomic charges be unambiguously fit to electrostatic potentials? Journal of Computational Chemistry. 1996; 17(3):367–383.

94. Lopes PEM, Lamoureux G, Mackerell AD. Polarizable empirical force field for nitrogen-containing heteroaromatic compounds based on the classical Drude oscillator. Journal of Computational Chemistry. 2009; 30(12):1821–1838. [PubMed: 19090564]

95. Harder E, Anisimov VM, Vorobyov IV, Lopes PEM, Noskov SY, MacKerell AD, et al. Atomic level anisotropy in the electrostatic modeling of lone pairs for a polarizable force field based on the classical Drude oscillator. Journal of Chemical Theory and Computation. 2006; 2(6):1587–1597.

96. Miller KJ. Additivity methods in molecular polarizability. Journal of the American Chemical Society. 1990; 112(23):8533–8542.

97. Baker CM, MacKerell AD. Polarizability rescaling and atom-based Thole scaling in the CHARMM Drude polarizable force field for ethers. Journal of Molecular Modeling. 2009; 16(3): 567–576. [PubMed: 19705172]

98. Yu HA, Whitfield TW, Harder E, Lamoureux G, Vorobyov I, Anisimov VM, et al. Simulating Monovalent and Divalent Ions in Aqueous Solution Using a Drude Polarizable Force Field. Journal of Chemical Theory and Computation. 2010; 6(3):774–786. [PubMed: 20300554]

99. Jorgensen WL, Madura JD, Swenson CJ. Optimized intermolecular potential functions for liquid hydrocarbons. Journal of the American Chemical Society. 1984; 106(22):6638–6646.

100. Jorgensen WL. Optimized intermolecular potential functions for liquid alcohols. The Journal of Physical Chemistry. 1986; 90(7):1276–1284.

101. MacKerell, AD. Atomistic models and force fields.. In: Becker, O., et al., editors. Computational Biochemistry and Biophysics. Marcel Dekker, Inc; New York: 2001. p. 7-38.

102. Yin D, MacKerell AD. Ab Initio Calculations on the Use of Helium and Neon as Probes of the van der Waals Surfaces of Molecules. The Journal of Physical Chemistry. 1996; 100(7):2588–2596.

103. Yin DX, MacKerell AD. Combined ab initio empirical approach for optimization of Lennard-Jones parameters. Journal of Computational Chemistry. 1998; 19(3):334–348.

104. Chen IJ, Yin D, MacKerell AD. Combined ab initio/empirical approach for optimization of Lennard-Jones parameters for polar-neutral compounds. Journal of Computational Chemistry. 2002; 23(2):199–213. [PubMed: 11924734]

105. Baker CM, Lopes PEM, Zhu X, Roux B, MacKerell AD. Accurate Calculation of Hydration Free Energies using Pair-Specific Lennard-Jones Parameters in the CHARMM Drude Polarizable Force Field. Journal of Chemical Theory and Computation. 2010; 6(4):1181–1198. [PubMed: 20401166]

106. Scott AP, Radom L. Harmonic Vibrational Frequencies: An Evaluation of Hartree-Fock, Møller-Plesset, Quadratic Configuration Interaction, Density Functional Theory, and Semiempirical Scale Factors. The Journal of Physical Chemistry. 1996; 100(41):16502–16513.

107. Pulay P, Fogarasi G, Pang F, Boggs JE. Systematic ab initio gradient calculation of molecular geometries, force constants, and dipole moment derivatives. Journal of the American Chemical Society. 1979; 101(10):2550–2560.
108. Foloppe N, Hartmann B, Nilsson L, MacKerell AD. Intrinsic conformational energetics associated with the glycosyl torsion in DNA: A quantum mechanical study. Biophysical Journal. 2002; 82(3):1554–1569. [PubMed: 11867468]
109. Foloppe N, Nilsson L, MacKerell AD. Ab initio conformational analysis of nucleic acid components: Intrinsic energetic contributions to nucleic acid structure and dynamics. Biopolymers. 2001; 61(1):61–76. [PubMed: 11891629]
110. Lin B, Lopes PEM, Roux B, MacKerell AD. Kirkwood-Buff analysis of aqueous N-methylacetamide and acetamide solutions modeled by the CHARMM additive and Drude polarizable force fields. Journal of Chemical Physics. 2013; 139(8):084509. [PubMed: 24007020]
111. Halkier A, Helgaker T, JÃ,rgensen P, Klopper W, Koch H, Olsen J, et al. Basis-set convergence in correlated calculations on Ne, N_2, and H_2O. Chemical Physics Letters. 1998; 286(3-4):243–252.
112. Graf J, Nguyen PH, Stock G, Schwalbe H. Structure and Dynamics of the Homologous Series of Alanine Peptides: A Joint Molecular Dynamics/NMR Study. Journal of the American Chemical Society. 2007; 129(5):1179–1189. [PubMed: 17263399]
113. Kirkpatrick S, Gelatt CD, Vecchi MP. Optimization by simulated annealing. Science. 1983; 220(4598):671–680. [PubMed: 17813860]
114. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of state calculations by fast computing machines. Journal of Chemical Physics. 1953; 21(6):1087–1092.
115. Shoemaker KR, Kim PS, York EJ, Stewart JM, Baldwin RL. Tests of the helix dipole model for stabilization of α-helices. Nature. 1987; 326(6113):563–567. [PubMed: 3561498]
116. Shoemaker KR, Kim PS, Brems DN, Marqusee S, York EJ, Chaiken IM, et al. Nature of the charged-group effect on the stability of the C-peptide helix. Proceedings of the National Academy of Sciences. 1985; 82(8):2349–2353.
117. Padmanabhan S, Marqusee S, Ridgeway T, Laue TM, Baldwin RL. Relative helix-forming tendencies of nonpolar amino acids. Nature. 1990; 344(6263):268–270. [PubMed: 2314462]
118. Fukunishi H, Watanabe O, Takada S. On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction. The Journal of Chemical Physics. 2002; 116(20):9058–9067.
119. Zhu X, Lopes PEM, Shim J, MacKerell AD. Intrinsic Energy Landscapes of Amino Acid Side-Chains. Journal of Chemical Information and Modeling. 2012; 52(6):1559–1572. [PubMed: 22582825]
120. Lopes PEM, Huang J, Shim J, Luo Y, Hui L, Roux B, et al. Polarizable Force Field for Peptides and Proteins based on the Classical Drude Oscillator. Journal of Chemical Theory and Computation, Articles ASAP (As Soon As Publishable) Publication Date (Web). Oct 17.2013 (Article) DOI: 10.1021/ct400781b.
121. Best RB, Buchete N-V, Hummer G. Are Current Molecular Dynamics Force Fields too Helical? Biophysical Journal. 2008; 95(1):L07–L09. [PubMed: 18456823]
122. Hegefeld WA, Chen S-E, DeLeon KY, Kuczera K, Jas GS. Helix Formation in a Pentapeptide: Experiment and Force-field Dependent Dynamics. The Journal of Physical Chemistry A. 2010; 114(47):12391–12402. [PubMed: 21058639]
123. Best RB, Mittal J, Feig M, MacKerell AD. Inclusion of Many-Body Effects in the Additive CHARMM Protein CMAP Potential Results in Enhanced Cooperativity ofα-Helix and β-Hairpin Formation. Biophysical Journal. 2012; 103(5):1045–1051. [PubMed: 23009854]
124. Lindorff-Larsen K, Maragakis P, Piana S, Eastwood MP, Dror RO, Shaw DE. Systematic Validation of Protein Force Fields against Experimental Data. Plos One. 2012; 7(2):e32131. [PubMed: 22384157]
125. Karplus M. Contact Electron-Spin Coupling of Nuclear Magnetic Moments. The Journal of Chemical Physics. 1959; 30(1):11–15.

126. Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. Journal of Physical Chemistry B. 2001; 105(28):6474–6487.

127. Oostenbrink C, Villa A, Mark AE, Van Gunsteren WF. A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. Journal Of Computational Chemistry. 2004; 25(13):1656–1676. [PubMed: 15264259]

128. Blanco FJ, Rivas G, Serrano L. A short linear peptide that folds into a native stable [beta]-hairpin in aqueous solution. Nat Struct Mol Biol. 1994; 1(9):584–590.

129. Muñoz V, Thompson PA, Hofrichter J, Eaton WA. Folding dynamics and mechanism of β-hairpin formation. Nature. 1997; 390(6656):196–199. [PubMed: 9367160]

130. Schuler B, Eaton WA. Protein folding studied by single-molecule FRET. Current Opinion in Structural Biology. 2008; 18(1):16–26. [PubMed: 18221865]

131. Jo S, Kim T, Iyer VG, Im W. CHARMM-GUI: A web-based graphical user interface for CHARMM. Journal of Computational Chemistry. 2008; 29(11):1859–1865. [PubMed: 18351591]

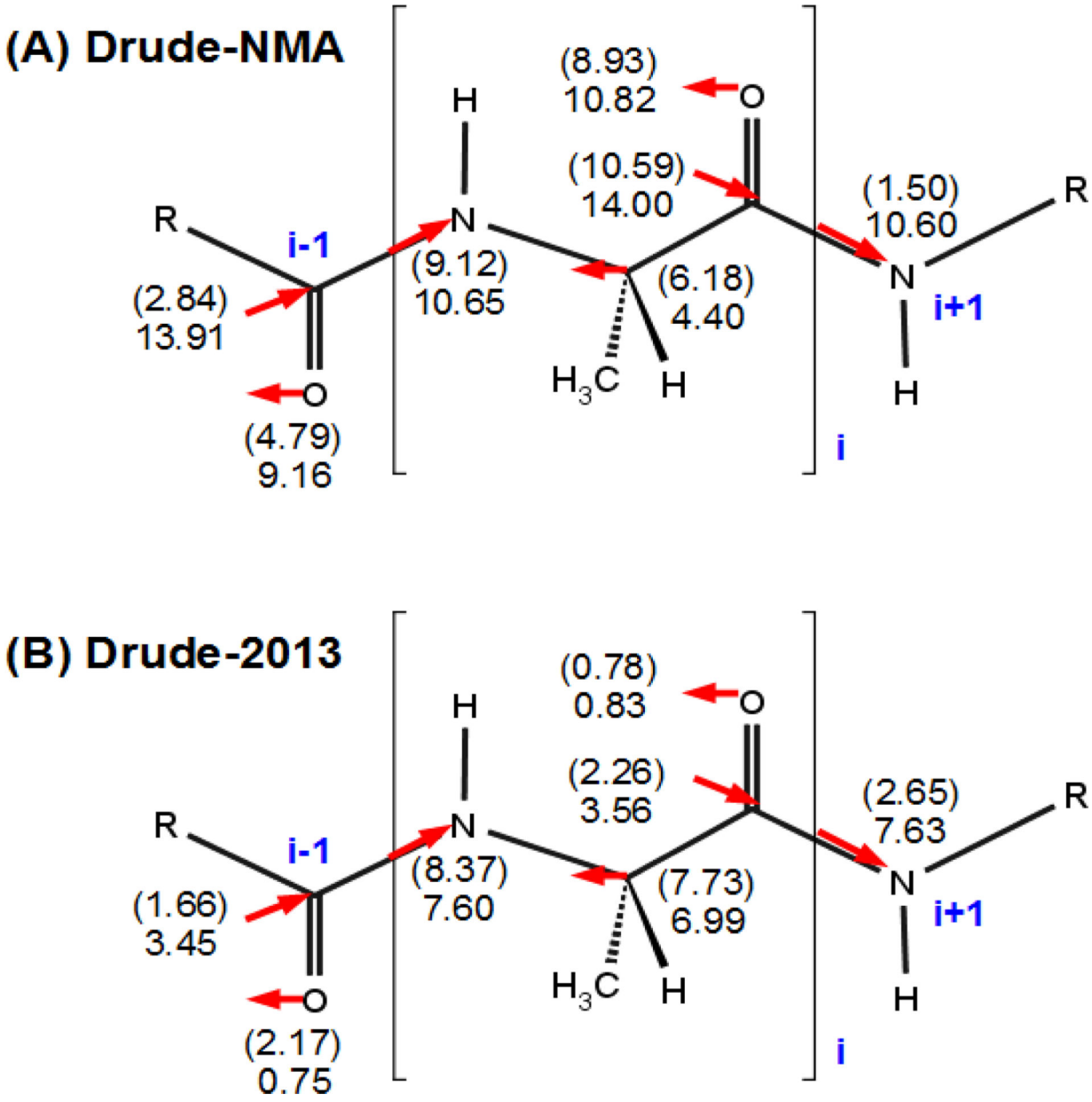

Figure 1.

Illustration of induced dipoles on dipeptide moieties of alanine dipeptide and $(Ala)_5$. Values in parenthesis are for alanine dipeptide.

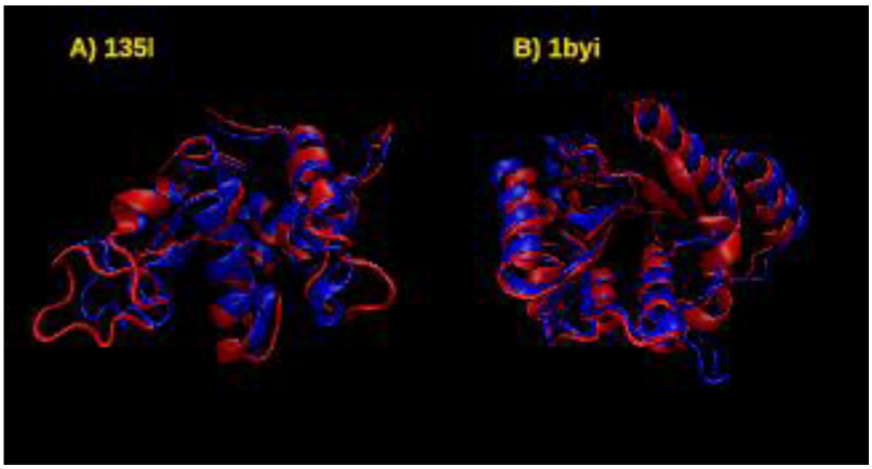

Figure 2.
100-ns snapshots from Drude-2013 simulations (red) of lysozyme (135L) and dethiobiotin synthase 1BYI superimposed on the starting crystallographic structures (blue).

Table 1

Gas phase dipole moments of alanine dipeptide and $(Ala)_5$[a], molecular polarizability of alanine dipeptide and relative energies of $(Ala)_5$.

Molecular dipole moment of alanine dipeptide (Debye)								
	αR				C5			
	QM[b]	Drude-NMA	Drude-ALA	Drude-2013	QM	Drude-NMA	Drude-ALA	Drude-2013
M	6.2	5.0	6.4	6.7	4.7	5.8	2.3	2.6
μ_x	1.3	0.1	3.1	3.0	−4.4	−5.6	−1.8	−2.3
μ_y	−1.6	−0.9	−1.7	−1.5	−1.0	−0.9	−0.6	−0.3
μ_z	5.9	4.9	5.4	5.8	1.2	0.9	1.3	1.3

Molecular dipole moment of $(Ala)_5$ (Debye)								
	αR				C5			
	QM[c]	Drude-NMA	Drude-ALA	Drude-2013	QM	Drude-NMA	Drude-ALA	Drude-2013
M	22.0	13.5	22.4	20.8	11.6	24.4	4.5	9.3

Molecular polarizability of alanine dipeptide (Å^3)								
	αR				C5			
	QM[b]	Drude-NMA	Drude-ALA	Drude-2013	QM	Drude-NMA	Drude-ALA	Drude-2013
A_{xx}	13.57	13.40	16.18	15.30	15.49	16.02	19.89	16.07
A_{yy}	12.72	12.60	14.29	14.36	12.06	11.87	13.39	12.78
A_{zz}	11.71	11.03	12.68	9.94	10.35	9.78	11.05	10.39

Relative energies of $(Ala)_5$ (kcal/mol)				
	QM[d]	Drude-NMA	Drude-ALA	Drude-2013
αR-C5	−6.59	6.21	5.31	−3.89
αR-PPII	−14.83	−5.77	0.42	−10.17

[a] $(Ala)_5$ is acetyl-$(Ala)_5$-N-methylamide

[b] QM dipole moments and polarizabilities of alanine dipeptide obtained at the B3LYP/aug-cc-pVDZ level with the polarizabilities scaled by 0.85

[c] QM dipole moments for $(Ala)_5$ obtained at the B3LYP/6-31G* level

[d] single point energies were calculated at the RIMP2/cc-pVTZ//RIMP2/cc-pVDZ level.