Published in final edited form as:
Curr Pharm Des. 2014 ; 20(20): 3281–3292.

Molecular Mechanics

Kenno Vanommeslaeghe[*], **Olgun Guvench**[†], and **Alexander D. MacKerell Jr.**[*]

Olgun Guvench: oguvench@une.edu; Alexander D. MacKerell: alex@outerbanks.umaryland.edu

[*]Department of Pharmaceutical Sciences, University of Maryland School of Pharmacy, 20 Penn St, HSF-II Rm 633, Baltimore, MD 21201; tel: 410-706-7442; fax: 410-706-5017

[†]Department of Pharmaceutical Sciences, University of New England College of Pharmacy, 716 Stevens Ave, Portland, ME 04103

Abstract

Molecular Mechanics (MM) force fields are the methods of choice for protein simulations, which are essential in the study of conformational flexibility. Given the importance of protein flexibility in drug binding, MM is involved in most if not all Computational Structure-Based Drug Discovery (CSBDD) projects. This section introduces the reader to the fundamentals of MM, with a special emphasis on how the target data used in the parametrization of force fields determine their strengths and weaknesses. Variations and recent developments such as polarizable force fields are discussed. The section ends with a brief overview of common force fields in CSBDD.

Keywords

Molecular Mechanics; Force Fields; Structure-Based Drug Design

1. Introduction

Computational Structure-Based Drug Discovery (CSBDD) involves drug design based on the three-dimensional structure of the biomolecular target, in most cases a protein. While the Quantum Chemical (better known as Quantum Mechanics or QM) methods discussed in section 2.2 are useful for studying the properties of isolated drug-like molecules and for limited studies on simple models of a protein's binding site, it is usually desirable to also perform simulations on the whole protein in the presence of water when applying CSBDD. Although some pioneering efforts have been made in applying semiempirical methods to this end,[1] doing so puts severe constraints on the simulation timescale, and semiempirical energy functions have weaknesses, such as poor treatment of dispersion interactions. Therefore, the method of choice for protein simulations remains Molecular Mechanics (MM) force fields, which approximate the quantum mechanical energy surface with a classical mechanical model, thereby decreasing the computational cost of simulations on large system by orders of magnitude. Furthermore, MM potential energy functions allow for a relatively accurate representation of dispersion interactions, which current QM methods only start recovering at the MP2 and higher levels of theory.[2]

2. Potential Energy Functions

This subsection will focus mainly on the class I additive potential energy function, which is the sum of bonded and nonbonded energy terms, as given by equation 1. This potential energy function, whose terms are described in Table 1, covers the vast majority of force fields used in CSBDD. Variations and recent developments in these models will be discussed below.

Bonded (intramolecular, internal), terms

$$E_{bonded}=\sum_{bonds}K_b(b-b_0)^2+\sum_{angles}K_\theta(\theta-\theta_0)^2+\sum_{\substack{improper\\dihedrals}}K_\varphi(\varphi-\varphi_0)^2+\sum_{dihedrals}\sum_{n=1}^{6}K_{\phi,n}(1+\cos(n\phi-\delta_n)) \quad (1)$$

Nonbonded (intermolecular, external) terms

$$E_{nonbonded}=\sum_{\substack{nonbonded\\pairs\,ij}}\frac{q_iq_j}{4\pi Dr_{ij}}+\sum_{\substack{nonbonded\\pairs\,ij}}\varepsilon_{ij}\left[\left(\frac{R_{\min,ij}}{r_{ij}}\right)^{12}-2\left(\frac{R_{\min,ij}}{r_{ij}}\right)^{6}\right]$$

2.1 Bonded Interactions

The class I potential energy function comprises 4 types of bonded interactions: bond stretching terms, angle bending terms, dihedral or torsional terms and improper dihedrals. As shown in equation 1, class I additive force fields approximate the bond stretching and angle bending contributions to the potential energy as harmonic oscillators as a function of the bond length and valence angle, respectively. In this approximation, only 2 parameters are needed for each bond and angle: the reference or equilibrium value (b_0 and θ_0) and the force constant (K_b and K_θ). The bond and angle terms dominate the local covalent structure around each atom and, in theory, when angle bending terms are present for all angles in a molecule, planar centers are kept planar by the sum of the reference angles θ_0 being 360° or higher so that any deviation from planar geometry would imply an increase in energy. However, there are cases where angular force constants K_θ that accurately reproduce the energetics of in-plane angular bending are not high enough to also reproduce the energetics of out-of-plane motions. Therefore, most if not all class I potential energy functions include an additional out-of-plane term, usually in the form of an improper dihedral, where the potential energy is harmonic as a function of the out-of-plane angle φ. Finally, the torsional energy is represented by a sum of cosine functions with multiplicities n=1,2,3… and amplitudes $K_{\varphi,n}$. The phases δ_n are usually constrained at 0° or 180° so that the energy surface of achiral molecules is symmetric and so that enantiomers have the same energetic properties. Note that typically, not all the cosine terms associated with each multiplicity are suitable for an accurate description of a given torsion; for example, only a 3-fold term (i.e. n=3) is appropriate for the H-C-C-H dihedral in ethane, while a 2-fold term is typically required for the treatment of double bonds such as the H-C=C-H dihedral in ethene. However, two or more multiplicities are often used for a given torsion angle to more

accurately treat the change in energy of the system as function of rotation about the central bond in the dihedral. In the end, the ability to correctly reproduce conformational energetics is an important criterion for the usefulness of a force field, and hence accurate parametrization of the dihedral terms is essential.

2.2 Anharmonicity and cross-terms

While the internal terms in Class I force fields are primarily harmonic or sinusoidal in nature, class II and III force fields contain cubic and/or quartic terms in the potential energy for bond and angles of the form $E_{bond} = K_b(b - b_0)^2 + K_b'(b - b_0)^3 + K_b''(b - b_0)^4 + \ldots$. While these higher-order terms allow for a more accurate reproduction of QM Potential Energy Surfaces (PES) and experimental properties such as vibrational spectra, they also introduce more parameters in the force field (K_b', K_b'',…), making optimization of the model more difficult. Moreover, Molecular Dynamics (MD) simulations associated with CSBDD are generally performed at room temperature, and the energy in bond and angle vibrations typically does not become high enough for anharmonicity to have a qualitatively important influence on the dynamics and energetics. Apart from anharmonic terms, class II and III force fields contain cross terms that reflect the coupling between adjacent bonds, angles and dihedrals. For example, a typical bond-bond cross term would be of the form $E_{cross}\ (b_1,b_2) = K_{b_1,b_2}(b_1 - b_{1,0})(b_2 - b_{2,0})$. Bond-angle, angle-angle, bond-torsion and angle-torsion cross terms can be introduced in a similar fashion. Some force fields (such as Allinger's MM2, MM3 and MM4) go even further by introducing cross terms that involve up to three internal coordinates (eg. bond-angle-bond and angle-torsion-angle).[3–8] A special case is the Urey-Bradley term, which consists of a harmonic potential as a function of the distance between the (non-bonded) atoms A and C of an A-B-C angle (i.e., as if there exists an extra bond stretching term between atoms A and C). This Urey-Bradley term is coupled with the A-B and B-C bond stretching terms and the A-B-C angle bending term through basic trigonometric relationships, and is a computationally elegant way of reproducing bond-bond coupling effects in vibrational spectra. However, compared to the more conventional cross-terms, it is poorly transferable to various combinations of 3 atoms with different reference values for the bonds or angle.

While anharmonicity and cross terms do allow for a better reproduction of subtle physical phenomena, they have the important disadvantage that their inclusion multiplies the amount of target data needed for meaningful optimization of the parameters. This dramatically increases the complexity of the parameter optimization process (see subsection 4). The higher target data requirement may not be a prohibitive hurdle for force fields that focus on reproducing the energetics of a limited number of small model compounds in vacuum, because large amounts of uniform and high-quality target data can be obtained through QM calculations. However, such an approach has proven inappropriate for the biomolecular force fields used in CSBDD, where nonbonded interactions and precise reproduction of the behavior of select torsions in the context of a large polymer in the condensed phase (eg. protein or nucleic acid backbone energetics) are vastly more important than the precise reproduction of bond and angle vibrations. For this reason, the biomolecular force field community has until recently refrained from introducing anharmonicity and cross terms, recognizing that there was still plenty of room for improvements within the framework of

the class I potential energy function. An exception is the recent introduction of the CMAP term in the CHARMM protein force field, which essentially is a torsion-torsion cross term consisting of a Ramachandran plot-like grid of correction energies that is a function of 2 dihedral angles.[9–12]

2.3 Nonbonded Interactions

As shown in equation 1, the electrostatics in class I force fields are simply handled by Coulomb interactions between fixed point charges q_i and q_j centered on the atoms, also known as “partial charges”. This treatment of electrostatic interactions is referred to as “additive” because the charges do not affect each other and all the individual atom-atom electrostatic interactions may simply be summed to yield the total electrostatic energy of the system. Beyond the partial atomic charge model, limited studies have been performed in which fixed dipole moments were associated with each atom,[13] or with extra point charges at fixed positions relative to some atoms.[14–17] It was found that as long as polarization was not included in these models, the improvements in intermolecular interactions for the elements typically present in bioorganic systems were usually not proportional to the greatly increased complexity in parametrization. An exception in this respect is water, for which 4-site models such as TIP4P[18], where a charge site is located along the H-O-H bisector, yield improved structural and dynamic properties as compared to 3-site models such as TIP3P[18] and SPC,[19] where partial charges are located only on the three atoms in the molecule.

For the van der Waals interaction component, a classical Lennard-Jones (LJ) 6-12 potential, defined by the radius $R_{\min,ij}$ and the well depth ε_{ij}, is typically used. The LJ potential is limited in its R^{-12} treatment of atomic repulsion, though again this limitation is not significant in CSBDD as the simulations are typically performed at room temperature. Some alternatives to the Lennard-Jones 6-12 have been proposed,[20–22] but with the exception of MMFF94’s “buffered-14-7” potential,[23, 24] these have not found widespread application in CSBDD because of increased computational expense, limited merit, and/or stability problems during MD simulations.

2.4 Polarization

The biggest disadvantage of the point charge model discussed above is that it does not allow for polarization. Indeed, molecules are known to have substantially higher dipole moments in the condensed phase than in the gas phase,[11] depending on the dielectric constant of the medium as well as other factors. Therefore, the charges used in a given force field are formally only valid in a given dielectric medium (i.e. one cannot perform condensed phase simulations with a gas phase force field or vice versa),[25] as the local dielectric constant inside, for example, a lipid bilayer or a protein is typically lower than in bulk water. To account for such differences additive force fields, even when highly optimized to take into account polarization in “the condensed phase,” will represent a compromise between the different types of dielectrics typically encountered in biomolecular systems. This problem, which is due to the lack of polarization in the electrostatic portion of the force field, is starting to become a limiting factor in the accuracy of additive force fields.[25–28] This has stimulated the development of polarizable force fields by several groups.[29–42]

There are three common schemes used for the introduction of polarization into an energy function.[28, 43, 44] These are the fluctuating charge model, the polarizable dipole (a.k.a. induced dipole) model, and the classical Drude oscillator (a.k.a. Shell or charge-on-spring (COS)) model. In the fluctuating charge model, the terms in the potential energy function are essentially the same as in non-polarizable force fields, except that the partial charges on the atoms in a molecule are allowed to redistribute in response to an external electric field. In the polarizable dipole model, every site *i* is given a dipole moment depending on its polarizability and the external electric field: $\boldsymbol{\mu}_{induced,i} = \alpha_i \mathbf{E}_i$. Finally, in the Drude oscillator model, a small part of every atom's total charge is represented by an extra point charge (i.e. the "Drude oscillator") that is bound to the parent atom by means of a harmonic force, the force constant of which is inversely proportional to the atom's polarizability. One variation of the model consists of giving the Drude oscillators their own Lennard-Jones potential as to simulate the Pauli repulsion of the electron clouds. The disadvantage of the latter variation is a significant increase in the number of parameters to be optimized, which is already high for polarizable force fields in general when compared to non-polarizable force fields.[26][45]

Regardless of the polarization scheme, it is important to note that when considering two interacting chemical entities A and B, the electrostatic potential around entity A will influence the polarization of entity B and hence the electrostatic potential around entity B. This in turn will influence the polarization of entity A and hence the electrostatic potential around entity A, and so forth... In theory, the most straightforward way to run useful calculations with polarizable force fields is to solve this polarization problem in a self-consistent fashion at every step of a simulation. However, doing so would incur an extreme computational performance hit. Therefore, extended Lagrangian methods have been developed,[46, 47] which involve treating the polarizable degrees of freedom as dynamic variables in the equations of motion. Such methods make it possible to run MD simulations using polarizable force fields at acceptable speeds, albeit more slowly than non-polarizable force fields.

Another important consequence of the cross-polarization described above it that the energy of a system can no longer be written as simple a sum of terms; in other words, the interaction energy for 3 or more interacting entities does not equal the sum of their pairwise interactions, as it does for non-polarizable, additive force fields. Such a limitation needs to be considered in CSBDD, where the energetic contributions of individual groups to binding are often calculated. While straightforward in an additive force field, this is not possible in polarizable models.

2.5 United-atom and coarse-grained models

A discussion of biomolecular force fields would not be complete without mentioning united atom and coarse-grained force fields. In the united atom model, nonpolar hydrogens are not explicitly represented by dedicated particles; instead, the Lennard-Jones parameters on the parent atom are optimized to include the (small) steric effect of the hydrogens. For polar hydrogens, the presence of a separate charged particle is necessary to represent hydrogen bonds and other polar interactions,[48] and polar hydrogens in united-atom force fields may be given a Lennard-Jones potential to prevent an infinite negative energy at the location of

the hydrogen atom.[49] In fact, initial incarnations of OPLS, AMBER, and CHARMM were united-atom force fields with polar hydrogens,[50–52] and current versions of GROMOS continue to be so.[25]

Coarse-grained models go one step further in that they represent whole chemical groups by a single entity, often called a “bead. ”[53–57] The number of atoms in one bead is a tradeoff between speed and accuracy, and varies from model to model. In the specific case of protein force fields, popular options include (but are not limited to) the “two-bead” models, where most amino acids are represented by two beads,[58–65] one for the backbone and one for the side chain, and the “four-bead” models, where the backbone is represented by three beads per residue and a fourth bead is used to represent the side chain.[66–70] Some coarse-grained models use ellipsoid beads, and the beads often have a dipole moment and sometimes higher multipole moments. Also, sophisticated functions are often used for bonded interactions, cross terms and hydrogen bonds, in order to recapture physical phenomena that were lost by aggregating atoms together in beads.

Reducing the number of particles by use of a coarse-grained model can greatly accelerate simulations, at the cost of accuracy. Coarse-grained models are especially popular for simulating very large biomolecular systems or studying phenomena that happen at very long time scales, such as protein folding.[71] However, it can be argued that coarse-grained models are of limited use for CSBDD because one typically would want to simulate a biomolecular system interacting with a drug-like molecule, and it would be difficult to design a set of “beads” that would allow simulating arbitrary drug-like molecules. Also, drug-target interactions often occur at the atomic level, and coarse-grained models may not be sufficiently detailed to accurately differentiate structurally similar drug candidates, such as those that are part of a congeneric series. Finally, as computers are ever getting more powerful, there is a general trend towards increased accuracy even if it incurs an increase in computational cost.

3. System preparation

The previous subsection discussed the mathematical functions that represent the energy in Molecular Mechanics methods. It was shown that these functions contain many parameters. The current subsection discusses how a typical Molecular Mechanics engine, when run by a user, determines which parameters to use for which atoms, bonds, angles,... in a molecule.

3.1 Atom typing

In the first stage, an atom type is assigned to each atom in the system under study. Different atom types are assigned to different elements, but also to different hybridization states of the same element, and usually even to atoms in the same hybridization state but a different chemical environment. For example, in the case of protein force fields, C_α would usually get a different atom type than the other sp^3 carbons, a carbonyl carbon would get a different atom type depending on whether it is part of an amide group or a carboxylate, there would be many different hydrogen types depending on the nature of the parent atom, and so forth. In the case of general force fields for organic molecules, atom types are assigned by rules that take into account the atom’s chemical environment. To fulfill this task, a number of

atom typers have been published including utilities in the Antechamber toolkit[72] associated with the General AMBER Force Field (GAFF),[73] the CGenFF program,[74, 75] MATCH[76] and SwissProt,[77] among others. In addition, most if not all commercial Molecular Mechanics packages include atom typing engines, although little information about their workings is in the public domain. It should be noted that more atom types make a force field more accurate, but also more difficult to parametrize and less transferable; this is discussed in more detail in subsection 3.3.

3.2 Charge assignment

Independently from the atom types, partial charges are assigned to each atom. In force fields for biopolymers, each monomer typically has a predetermined set of charges (as well as atom types) which are provided as part of the force field and which can readily be applied to a protein or nucleic acid. Typically, these monomer charge distributions were initially assigned using QM target data (see subsection 4.2.5), then further optimized to reproduce X-ray structures (subsection 4.2.1), dielectric constants (subsection 4.2.4) and/or thermodynamic properties (subsection 4.2.6). Although these elaborate charge optimization approaches guarantee high-quality charges, such methods are not practical for large numbers of drug-like molecules. Instead, most force fields for organic molecules include charge assignment schemes that are applied “on-the-fly” by the modeling software whenever the user starts a calculation on a new molecule. Usually, these are either bond charge increment methods, electronegativity equalization methods, or a combination thereof. A pure bond charge increment scheme, like the one implemented in MMFF94,[23, 24, 78] would start with assigning initial “formal” charges, which would be 0 for all atoms except the ones that belong to a chemical group having a net charge. In the next step, for every bond in the molecule, an amount of charge would be transferred between the bonded partners depending on their atom types.[79] Bond charge increments for every combination of atom types would typically be read from a predetermined table that is stored along with the other force field parameters (see subsection 3.3 below). A more recent example of a force field that uses a variation of the bond charge increment methodology for assigning charges on molecules that were not explicitly parametrized is the CHARMM General Force Field (CGenFF).[75, 80] Electronegativity equalization schemes come in many flavors, but the basic principle is that every atom type has a predetermined “inherent electronegativity” and “chemical hardness”, and the “instantaneous” electronegativity of an atom *i* in a molecule is given by $\chi_i=\chi_i^o+2\eta_i q_i$, where χ_i^o is atom *i*’s inherent electronegativity, η_i its chemical hardness and q_i is the partial charge on the atom. Initial charges are set in a similar fashion as for the bond charge increment methods, then charge is redistributed between the atoms in an iterative fashion until all atoms have the same instantaneous electronegativity. Note that the basic electronegativity equalization scheme as described above does not yield chemically sensible results, and practical electronegativity equalization methods are significantly more complicated;[81–83] the specific details of these schemes are beyond the scope of this chapter. An alternative approach that has become practical in recent years is to do a quick QM calculation on the molecule of interest prior to the start of an MM simulation, and derive partial charges from the QM wave function. However, popular charge assignment schemes that directly perform a per-atom partitioning of the electron density (eg. Mulliken, NPA, AIM, Hirshfeld aka. stockholder,…) generally fail to reproduce the Electrostatic

Potential (ESP) around the molecule when translated to atom-centered point charges, and are hence unsuitable for calculating electrostatic interactions. One workaround for this problem is to apply empirical bond charge corrections to the charges that result from partitioning. Since these empirical corrections can in theory be parametrized to correct for multiple approximations at the same time, they can be combined with a low (often semi-empirical) level of theory and an inexpensive (often Mulliken) partitioning scheme. One prominent example of this approach is AM1-BCC,[84] which can be used with the AMBER force field[85] and GAFF.[72] More sophisticated empirical correction schemes are used in Cramer, Truhlar *et al*'s CM1-CM5 models;[86–88] CM1 and CM3 have been validated for use with the OPLS force field.[89] Conversely, a non-empirical way to derive relevant point charges from the electron density is to consider the QM ESP on a large number of grid points around the molecule and optimize the partial charges in an automated fashion to reproduce this electrostatic potential as closely as possible.[90–92] To prevent this methodology from producing unphysical charges, in particular on buried atoms, restraints are often built into the charge optimization procedure. A widely used example of this approach is the RESP charge fitting procedure[93] as implemented in the Antechamber toolkit[72] that assigns charges for use with GAFF.[73] None of these charging schemes are perfect,[94] and assignment of appropriate charges is still considered one of the big challenges in small molecule force fields.

3.3 Parameter assignment

After assigning atom types and partial charges q_i, the other parameters in equation 1 are read from predetermined tables based on the atom types. Each atom type typically carries its own ε_i and $R_{\mathrm{min},i}$ value, which are usually averaged between two atoms i and j to obtain the ε_{ij} and $R_{\mathrm{min},ij}$ to be used to represent the LJ interaction between these atoms. However, some biomolecular force fields (most notably OPLS-AA and GROMOS; see subsection 5) use the geometric mean for both quantities, while others (most notably AMBER and CHARMM) use the Lorentz-Berthelot combining rules, which consist of using the arithmetic mean for R_{min} and the geometric mean for ε. This difference in combining rules (a.k.a. mixing rules) is one of the reasons why LJ parameters cannot be transferred between force fields. Given the fact that torsional parameters in many cases serve as compensation for imperfections in the nonbonded interactions (see subsection 2.1), this implies that torsional parameters also cannot be transferred between force fields. In essence, the charges, LJ and torsional parameters in a given force field are largely dependent on one another. This is part of the rationale for the often-repeated statement that different force fields are usually not compatible. Finally, it should be noted that the nonbonded interactions between 1–2 and 1–3 atom pairs (i.e. covalently bound to each other for 1–2 or separated by 2 covalent bonds for 1–3) are usually excluded from the potential energy function. Additionally, some force fields ignore 1-4 nonbonded interactions, scale all 1-4 interactions by a predetermined factor, or have a second ε and R_{min} value for the 1-4 interactions associated with specific atom types. Regardless of these nonbonded exclusion schemes, 1-4 and longer-range intramolecular nonbonded interactions vary in energy with the dihedral angles in a molecule, giving rise to a nonbonded contribution to the dihedral PES. This has important implications for dihedral parameter optimization; see subsection 4.1.

As for the bonded parameters, these are usually simply read from tables that contain a reference distance b_0 and force constant K_b for every combination of 2 atoms that might form a bond, a reference angle θ_0 and force constant K_θ for combinations of 3 atoms, and dihedral and improper parameters ($K_{\varphi,n}$, δ_n, K_ϕ, ϕ_0) for combinations of 4 atoms. In practice, this implies that the number of bonded parameters in a general organic force field is roughly proportional to the cube of the number of atom types. Some force fields alleviate this combinatorial explosion by providing wildcard parameters for which only the two inner atoms of a dihedral are specified; the wildcards are used when no dihedral parameters are available for a specific combination of 4 atom types.

As mentioned in subsection 3.1, increasing the specificity of the atom types (and thus their number) in a force field can in principle improve its accuracy by allowing for more specific parameters for slightly different chemical groups. On the other hand, the number of atom types should be limited in order to keep the parametrization feasible; having the ability to differentiate subtly different chemical groups is of no help if a sufficient number of parameters to cover all these functional groups cannot be optimized in a reasonable amount of time. Moreover, having more atom types in a force field increases the chance that a user's molecule of interest contains a combination of atom types that was not considered during the force field's design, thus decreasing the transferability of the force field from explicitly parametrized chemical groups to novel moieties. In summary, the number of atom types is a tradeoff between accuracy on the one hand and feasibility and transferability on the other hand.

4. Parameter Optimization

As discussed in the previous sections, the two main elements of a force field are its potential energy function and the set of parameters used in that energy function. It can be argued that to some extent, the parameters are more important than the potential energy function in defining the scope and quality of a force field; even the simple class I potential energy function can produce very accurate results if combined with well-optimized parameters, while even the most sophisticated potential energy function will fail to produce meaningful results when given unreasonable parameters, or may even be unstable for the purpose of running MD simulations. In this subsection, a short overview of the general procedure for determining these parameters (a.k.a. "parameter optimization" or simply "parametrization") is given. Subsequently, the target data for parameter optimization are discussed in detail, because ultimately, the scope and quality of a set of parameters is largely dependent on the nature of these target data.

4.1 Optimization Procedure

The procedure for parametrizing force fields is conceptually very simple. First, a set of model compounds is constructed, consisting of small molecules containing the chemical groups of interest. Then, target data for these model compounds are collected, which is discussed in detail in subsection 4.2 below. Large amounts of target data are needed to avoid an algebraically underdetermined parametrization (see below), and the choice of model compounds is often inspired by the availability of target data. In the next phase, all parameters needed for a full force field-based description of the model compounds are set to

so-called “initial guess” values. In principle, these initial guesses are not critical because they will be optimized in the next steps, but in practice, a good initial guess can greatly facilitate the optimization. Then, the following steps are performed:

- calculation of the target data using the force field
- comparison of the force field results with the actual target data
- adjustment of the force field parameters

These three steps are executed repeatedly until the force field reproduces the target data satisfactorily. However, adjusting all the parameters and recalculating all the target data at each iteration would be impractical. Therefore, the different parts of the force field such as the charges, the equilibrium values, the force constants, and the dihedral parameters, are typically optimized separately. This sequence of partial optimizations, each of which is iterative, should in turn be repeated several times until no more adjustment of any class of parameters is needed. However, in practice, optimizing the different parts in a sensible order may limit the number of iterations to 2 or 3. Specifically, the bonded parameter space includes “hard” degrees of freedom (bond, angle and improper dihedral parameters) that are characterized by harmonic functions with relatively high force constants, and “soft” degrees of freedom (mostly dihedrals) that are responsible for large conformational changes. It is usually advantageous to optimize parameters in a hard-to-soft order because the hard degrees of freedom influence the softer ones much more than *vice versa*. For example in biphenyl, the reference length and force constant of the central bond as well as the C-C-H angle force constant strongly modulate the amount of H…H steric clash and electrostatic repulsion that will occur in the planar conformation, and hence the rotational barrier. Thus, depending on small variations in these hard degrees of freedom, widely different values for the dihedral parameter around the central bond may be required in order for the sum of the nonbonded and bonded dihedral PES to reproduce a target (experimental or QM; see subsection 4.2) torsional barrier. The intramolecular electrostatic interactions that are part of the dihedral PES are also strongly dependent on the charges, which would appear to provide an incentive for optimizing the charges before the bonded parameters. However, most common charge optimization schemes are performed on a single conformation that is relevant for simulations; often an MM minimized conformation is selected for this purpose. The location of this minimum is in turn dependent on the bonded parameters. The authors propose to work around this by first optimizing the charges on a QM minimized conformation, then optimizing the bonded parameters in a hard-to-soft order (which will make the MM minimum approach the QM minimum), and finally re-optimizing the charges on an MM minimized conformation.[80]

The iterative nature of the force field optimization procedure has a number of important consequences:

- The parameter set is essentially a large vector consisting of scalar quantities that are *a priori* unknown, and the collection of target data is a large vector of known scalar quantities that can be reproduced by applying a (usually nonlinear) mathematical function on the parameter set vector. Thus, parameter optimization is formally equivalent to solving a (very complex) system of equations. This implies that if

there are less target data points than parameters, the parametrization is algebraically underdetermined, as more than one different combination of parameters can yield the same values for the target data. At this point, it should be noted that most real-life force fields are underdetermined to some extent. This is not necessarily a problem as long as common sense is applied during the parametrization so that the parameters have physical values. Indeed, with current force fields, there are enough parameters available to make it straightforward to determine whether an optimized parameter is physically meaningful as well as consistent with the remainder of the force field. This property is more formally exploited in the incremental or "build up and transfer" parametrization procedure, where the smallest possible model compound containing a functional group of interest (eg. methanol in the case of alcohols) is first parameterized, followed by stepwise larger molecules (eg. ethanol, propanol, isopropanol). At each incremental step, the parameters from the smaller model compound are retained (insofar that they are in acceptable agreement with the target data), and only the newly introduced parameters in the larger molecule are optimized. Nevertheless, if the "undeterminedness" is too high during the parametrization of any part of the force field, optimizing the parameters becomes a difficult task. More concrete examples of this problem are given in the discussion of the target data below.

- The quality of the force field is critically dependent on the target data. Any physical phenomenon (e.g. condensed phase effects) that is not accounted for in the target data cannot be expected to be reproduced accurately by the force field.
- The target data should ideally be easy and fast to reproduce using force field methods. Indeed, the phrase "feasible for the purpose of force field parametrization" in the discussion of the target data below essentially means: "feasible to be repeated many times for many different model compounds and parameter sets".

4.2 Target Data

In principle, any property of any system in which a relevant model compound is present can be used as target data for parameter optimization. However, there are criteria that make some types of data more appropriate or convenient than others. Ideally, target data for force field parametrization should be:

- representative of the chemical environment in which the final force field will operate,
- easy to correlate with one parameter, if not a small number of parameters,
- computationally inexpensive to reproduce with the force field, as this is repeated many times during parameter optimization,
- plentiful, otherwise the force field might be algebraically underdetermined (see subsection 4.1 above),
- precise

In practice, some of these criteria turn out to be mutually exclusive to some degree, and target data rarely satisfy all of them. In the following sections, a few common sources of target data are discussed in more detail along with their advantages and disadvantages. Specifically, we will focus on small molecule X-ray crystallography, IR and NMR spectroscopy, dielectric constants, QM target data and thermodynamic properties. It should be noted that this list is intended to be representative but not exhaustive; for example, neutron and X-ray scattering data have been used for specialized purposes such as water and lipid force field parametrization.[95–98]

4.2.1 Small molecule X-ray crystallography—Small molecule crystal geometries are a good source of reference values for the “hard” degrees of freedom (i.e. the bonds and angles). Individual crystals may contain packing artifacts or may simply be biased because they represent only one conformation, but these deviations tend to cancel out when averaging values over many different crystal structures. Surveys of large collections of crystal structures may even provide hints about relative populations of different conformations and thereby can serve as target data for dihedral optimization. Before sufficiently high-quality QM methods became computationally accessible for the purpose of force field optimization, most of the target data for optimizing bond and angle parameters came from small molecule crystals.[50, 99, 100] Finally, simulating existing crystals of small organic molecules, short peptides, or even whole proteins if the experimental data is of adequate resolution, is a good way to validate and fine-tune the force field as a whole.[10, 101, 102]

4.2.2 IR spectroscopy—Crystal geometries provide plenty of useful target data for optimization of reference or equilibrium parameters, but are of limited use when optimizing force constants. For this purpose, the most straightforward experimental source is IR spectroscopy.[103, 104] However, in a typical IR spectrum, only a limited number of signals can be unambiguously correlated to specific bond stretching or angle bending motions. Indeed, each peak in an IR spectrum represents one of the normal modes of the whole molecule, and even if an experimental normal mode consists predominantly of motions of one particular bond or angle, the composition of the modes might be different in the force field, and it may not be trivial to use all available IR spectroscopic data for force field optimization. Lastly, the problem of algebraic underdeterminedness often makes it more difficult to optimize force constants using vibrational target data, and typically, many assumptions have to be made during the process. To overcome this, vibrational spectra from QM methods are of great utility, as discussed in subsection 4.2.5 below.

4.2.3 NMR spectroscopy—Although statistical data from crystal surveys can successfully be used for optimizing dihedral parameters, a large number of crystal structures containing the dihedral of interest is needed for the statistics to be sufficiently reliable, and even then, bias related to the crystal environment or the selection of compounds that were crystallized can never be ruled out. Therefore, independent validation of the resulting dihedral parameters is often desired. Conveniently, coupling constant from simple 1D NMR experiments can be correlated with a singe dihedral’s conformational preferences through the Karplus equation and variants thereof.[105–109] The most correct way of utilizing this

target data consists of a solution-phase MD simulation followed by prediction of the coupling constant using this equation. An example of the use of coupling constants for force field optimization and validation are studies on proteins and carbohydrates from our laboratory.[12, 110–112] Similarly, the proximity between atoms during an MD simulation can be correlated to data from Nuclear Overhauser Effect (NOE)-based NMR techniques such as NOESY and ROESY for the purpose of validating and fine-tuning force fields.[113] Additionally, NOE data can be used as target data in a similar fashion as crystallographic data[114–121] (see subsection 4.2.1). Finally, spin-relaxation data can be correlated to the effective correlation time in simulations.[122] However, reproducing NMR target data remains a moderately burdensome process and the data itself is often somewhat heterogeneous.

4.2.4 Dielectric constants—Dielectric constants, which directly correlate with a molecule's dipole moment and therefore its partial atomic charges, can be reproduced using MD simulations of pure solvents.[123–125] Such approaches are computationally demanding, requiring simulation times of tens of nanoseconds. However, when performing high-level optimization of nonbonded parameters on molecules for which dielectric data are available such calculations are warranted. Efforts towards the development of polarizable force fields have extensively used reproduction of dielectric constants as target data, as the dielectric constant is dependent on the polarizability of the model as well as the dipole moment.[35, 126, 127]

4.2.5 QM target data—As it is usually problematic to collect enough experimental target data for optimization of all aspects of a force field, results from QM calculations are often used as target data. Methods and models from the field of theoretical chemistry have been used for charge assignment since the early days of molecular mechanics, the electronegativity equalization schemes discussed in subsection 3.2 being a simple example. More recently, methods such as RESP,[93] based on full electronic structure calculations, gained prominence. A competing approach for force fields aimed at simulations in aqueous solution is to construct a system consisting of a model compound and a water molecule, optimize the distance between these two partners at the QM level, and calculate the QM interaction energy. If the resulting water interaction distance and energy are properly scaled to take into account condensed phase effects, they make excellent target data for charge optimization, especially if the procedure is repeated with multiple water molecules at different positions around the model compound.[48, 50, 128] The main disadvantage of this approach is that it is difficult to automate, so the user cannot automatically compute charges on their compound of interest prior to simulation. However, as discussed in subsection 3.2, this is not an issue for biopolymers, where predetermined charges for the monomers are used. Indeed, this approach was applied successfully for charge assignment in the OPLS and CHARMM biomolecular force fields.[50, 80, 128, 129]

QM methods are also of great utility for the production of target data for bonded parameter optimization, often supplanting the use of crystal data and IR spectroscopy.[80] This is largely due to increases in computer power and advances in computational methods making explicit treatment of electron correlation in QM calculations feasible, allowing for improved

treatment of dispersion (though limitations in the treatment of dispersion still exist). Compared to crystal surveys, QM reference bond lengths and angles are more easy to obtain, especially for unusual species, and more homogeneous than crystal data if calculated within a carefully designed protocol. The results of a QM vibrational analysis can be correlated to force constants in the force field more easily than peaks in an experimental spectrum, owing to the fact that the Potential Energy Distribution (PED) of each normal mode is known. Finally, dihedral potential energy scans can routinely be applied on small and medium-sized model compounds, yielding ideal target data for parametrization of dihedral terms.[130–132] The only disadvantage of the use of QM methods is that, due to the fact that explicit solvent QM simulations are too costly and current implicit solvent methods are too unreliable, QM target data are usually determined in vacuum. Although offsets and scaling factors have been proposed to approximately take into account condensed phase effects,[48, 50, 128] additional validation and refinement of QM-based force fields against the abovementioned sources of experimental target data are of great importance.

4.2.6 Thermodynamic properties—The most difficult term in current force fields to optimize are the LJ parameters. While it is possible to use crystal geometries and QM interactions with water molecules[129] and noble gas atoms[133] as sources of target data for this purpose, it was found that correct van der Waals behavior is critically dependent on condensed phase and dynamic effects, and exclusive use of crystallographic and QM data is unreliable.[134] A more accurate way to optimize LJ parameters, pioneered by Jorgensen, is to use experimental liquid densities and heats of vaporization of small model compounds as primary target data, optionally in combination with other thermodynamic data and with guidance from QM data, especially with respect to the *relative* values of LJ parameters.[50, 133, 135, 136] Liquid densities and heats of vaporization can be reproduced relatively easily through Monte Carlo (MC) or short MD simulations, and it was found that LJ parameters obtained in this way exhibit a high degree of transferability towards atoms of the same element and hybridization state. Indeed, force fields that were parametrized based on this principle often employ the same LJ parameters for a number of different atom types. One limit of this approach is that it cannot be applied to species that have a net charge or a decomposition temperature that is lower than their melting point, as these do not exist as a bulk solvent. In these cases, crystal lattice parameters, heats of sublimation and free energies of solvation can be used, which are available for a range of molecules, including charged species and compounds that are not liquid at room temperature. However, complicated and computationally expensive free energy perturbation or thermodynamic integration methods are needed to reproduce free energies of solvation, and this has only recently became feasible for the purpose of force field optimization.[126, 137–140]

5. Common Force Fields in Structure-Based Drug Discovery

As touched upon earlier in this chapter, each force field has a scope. This is not only true in the chemical sense, i.e. by targeting a specific range of chemical groups or class of molecules, but also in a wider sense: force fields can be optimized for performing MD simulations or more static calculations such as energy minimizations,[141] they can be parametrized towards reproducing the conformational behavior of single molecules or

towards reproducing intermolecular interactions, and they can be specialized for gas phase calculations or for condensed phase simulations.

A primary requirement for force fields for use in CSBDD is that they must support biomolecules, or at the very least proteins, in a biologically relevant environment. Such environments are typically aqueous solution, although support for lipid bilayers is desirable, especially if transmembrane proteins such as G-protein coupled receptors (GPCR) are being considered. Also, eligible force fields must support drug-like molecules in solution phase and should be explicitly parametrized to reproduce intermolecular interactions. These requirements limit the utility of the majority of general force fields for organic molecules such as UFF,[142] CFF93,[8] and Allinger's MM2,[4] MM3[3, 5, 6] and MM4[7] force fields. Indeed, although some of these force fields are very capable for and widely used in Computational *Ligand*-Based Drug Design (CLBDD), they are parametrized mainly using gas-phase target data and (in some cases) condensed-phase target data that are not directly related to intermolecular interactions, and may therefore be assumed to be unsuitable for condensed phase studies of protein-drug interactions. An exception is MMFF94,[23] which was created with the express goal of enabling solution-phase simulations on ligand-biomolecule systems and obtaining accurate interaction energies. A pioneering effort when it was published over fifteen years ago, it was the most capable and generally applicable force field for CLBDD in existence, and its utility was significantly bolstered by the associated atom typing engine that was able to automatically assign parameters and rate their qualities for arbitrary drug-like small molecules. Nevertheless, a limitation of MMFF94, especially for the purpose of CSBDD, was that optimization of its nonbonded parameters was mainly driven by reproduction of QM interaction geometries and energies, and it was subsequently shown that using thermodynamic target data (see subsection 4.2.6), though computationally burdensome (particularly at that time in history) yielded more reliable nonbonded interactions for the condensed phase.[134] The importance of basing nonbonded parameters on the use of condensed phase data can be explained by the fact that nonbonded interactions in additive force fields are summed over atom pairs, while in reality, multi-body interactions contribute significantly to a system's condensed phase behavior.[125, 143, 144] In this sense, nonbonded parameters that are optimized using condensed phase properties contain an implicit correction for multi-body interactions and are therefore often referred to as "effective pairwise potentials." A more fundamental limitation of general force fields is that accurately reproducing the subtle interactions in biomolecules requires a dedicated biomolecular force field; covering a wide chemical space inherently compromises accuracy in representing classes of molecules in a well-defined chemical space, such as proteins. Like MMFF94, the commercial Tripos force field,[145] CVFF,[146] and Momany and Rone's commercial CHARMm force field[147] (not to be confused with the academic CHARMM force field; note the different capitalization) are general force fields that aim to provide coverage for proteins, parametrized using similar target data, and thus prone to similar limitations.

This brings us to the topic of biomolecular force fields, which are highly optimized for solution phase MD simulations on biomolecules. In this subsection, we will focus on the widely-used biomolecular protein force fields AMBER[148], CHARMM[12, 102],

GROMOS [25], and OPLS-AA [149], which account for the majority of recently published MD simulations of proteins.[150] All four of these force fields have been developed by academic research groups and, accordingly, the associated parameters have been peer-reviewed and made publicly available. Though we limit our discussion to these four, we do note that the ENCAD force field[151, 152] has been used extensively by Daggett and coworkers to study protein folding via explicit solvent MD simulations [153], and the ECEPP[154–158] force field of Scheraga and coworkers has been used for protein structure prediction with an implicit solvent treatment.[159–161]

While it is important to use a well-validated biomolecular force field in CSBDD, in most cases, practical simulations should also include a drug candidate. Although it is tempting to represent the biological part of a drug-target system by a biomolecular force field and the organic part (i.e. the drug) by an organic force field, it is unlikely that this will yield properly balanced intermolecular interactions, because the nonbonded parameters in different force fields are developed using different strategies and combining rules (see subsection 3.3). To overcome this problem, efforts were started to create organic force fields that are specifically meant to be used with existing, highly optimized and tested biomolecular force fields. As a result of this effort, AMBER[148, 162] now includes the General AMBER Force Field (GAFF)[73] and the Antechamber toolkit,[72] which allow the user to generate an AMBER force field model for an arbitrary input molecule. OPLS-AA,[163–165] whose optimizations emphasized condensed phase properties of small molecules, has been extended to cover a diverse set of small molecule model compounds and may be a good choice, though atom type assignment must be done by hand (although there exists a commercial implementation of OPLS-AA with atom typing functionality[166]). CHARMM has been extended with the CHARMM General Force Field (CGenFF), which covers a wide range of chemical groups present in biomolecules and drug-like molecules including a large number of heterocyclic scaffolds,[80, 113] and a web interface for automatic atom typing and assignment of parameters and charges by analogy has recently been published.[74, 75] Finally, the GROMOS force field atom type palette, which derives from parameters for biopolymers, also provides a reasonable amount of diversity for the construction of force field models of small molecules.[25] While varying ranges of drug-like organic molecules are covered by these force fields, they typically represent extrapolations of those force fields. Accordingly, the issue of transferability of the force field becomes important, such that the more significant the extrapolation from chemical moieties explicitly optimized in the force field, the less reliable the resulting parameters. For this reason, tools such as the CGenFF parameter and charge assignment interface return penalty scores that roughly reflect the degree of extrapolation. The general non-transferability of force fields also makes it necessary for the user to validate the parameters using one or more of the methods discussed in subsection 4 above.

6. Summary

Molecular mechanics based theoretical methods are widely applied to CSBDD. Central to the utility of empirical force fields in CBSDD is computational efficiency due to the simple form of the potential energy function, while at the same time achieving a suitable level of accuracy as well as coverage. Suitable for CBSDD are those force fields highly optimized

for biomolecular systems and then extended to drug-like compounds. However, the transferability of molecular mechanics parameters must always be considered and appropriate tests of the model undertaken when applied to new molecules. Concerning the future, improvements in the range of molecular entities covered by additive force fields can be expected as well as more automated methods for force field validation and optimization. In addition, over the next several years it is anticipated that models that include the explicit treatment of polarization will be become available. While these will initially be limited to biomolecules, they will certainly be extended to drug-like molecules, offering the potential for improvements in the accuracy of empirical force fields in the context of CSBDD.

Acknowledgments

Financial support from the NIH (GM051505, GM070855, GM0772558 and GM099022), the NSF (CHE-0823198), the University of Maryland Computer-Aided Drug Design Center, and University of New England start-up funds is acknowledged.

References

1. Liu H, Elstner M, Kaxiras E, Frauenheim T, Hermans J, Yang W. Quantum mechanics simulation of protein dynamics on long timescale. Proteins: Structure, Function, and Genetics. 2001; 44:484–9.
2. Hobza P, Sponer J. Structure, Energetics and Dynamics of the Nucleic Acid Base Pairs: Nonempirical *Ab initio* Calculations. Chemical Reviews. 1999; 99:3247–76. [PubMed: 11749516]
3. Lii J-H, Allinger NL. Molecular Mechanics. The MM3 Force Field for Hydrocarbons. 3. The van der Waals' Potentials and Crystal Data for Aliphatic and Aromatic Hydrocarbons. Journal of the American Chemical Society. 1989; 111:8576–82.
4. Allinger NL. Conformational Analysis. 130. MM2. A Hydrocarbon Force Field Utilizing V1 and V2 Torsional Terms. Journal of the American Chemical Society. 1977; 99(25):8127–34.
5. Allinger NL, Yuh YH, Lii J-H. Molecular Mechanics. The MM3 Force Field for Hydrocarbons. 1. Journal of the American Chemical Society. 1989; 111:8551–66.
6. Lii J-H, Allinger NL. Molecular Mechanics. The MM3 Force Field for Hydrocarbons. 2. Vibrational Frequencies and Thermodynamics. Journal of the American Chemical Society. 1989; 111:8566–75.
7. Allinger NL, Chen KH, Lii JH, Durkin KA. Alcohols, ethers, carbohydrates, and related compounds. I. The MM4 force field for simple compounds. J Comput Chem. 2003 Sep; 24(12): 1447–72. [PubMed: 12868110]
8. Hwang MJ, Stockfisch TP, Hagler AT. Derivation of Class II Force Fields. 2. Derivation and Characterization of a Class II Force Field, CFF93, for the Alkyl Functional Group and Alkane Molecules. Journal of the American Chemical Society. 1994; 116:2515–25.
9. MacKerell AD Jr, Feig M, Brooks CL III. Accurate treatment of protein backbone conformational energetics in empirical force fields. J Am Chem Soc. 2004; 126:698–9. [PubMed: 14733527]
10. MacKerell AD Jr, Feig M, Brooks CL 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; 25:1400–15. [PubMed: 15185334]
11. MacKerell AD Jr. Empirical Force Fields for Biological Macromolecules: Overview and Issues. J Comput Chem. 2004; 25:1584–604. [PubMed: 15264253]
12. 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): 3257–73. [PubMed: 23341755]
13. Price SL, Harrison RJ, Guest MF. An Ab Initio Distributed Multipole Study of the Electrostatic Potential Around an Undecapeptide Cyclosporin Derivative and a Comparison with Point Charge Electrostatic Models. J Comput Chem. 1989; 10:552–67.

14. Vinter JG. Extended Electron Distributions Applied to the Molecular Mechanics of Some Intermolecular Interactions. Journal of Computer-Aided Molecular Design. 1994; 8:653–68. [PubMed: 7738602]
15. Hunter CA. Sequence-dependent DNA Structure. The role of Base Stacking Interactions. Journal of Molecular Biology. 1993; 230(3):1025–54. [PubMed: 8478917]
16. Packer MJ, PDM, Hunter CA. Sequence-dependent DNA structure: Dinucleotide Conformational Maps. Journal of Molecular Biology. 2000; 295:71–83. [PubMed: 10623509]
17. Hunter CA, Saunders KM. The Nature of π-π interactions. Journal of the American Chemical Society. 1990; 112:5525–34.
18. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of Simple Potential Functions for Simulating Liquid Water. Journal of Chemical Physics. 1983; 79:926–35.
19. Berendsen, HJC.; Postma, JPM.; van Gunsteren, WF.; Hermans, J. Interaction Models for Water in Relation to Protein Hydration. In: Pullman, B., editor. Intermolecular Forces. Dordrecht, Holland: Reidel Publishing Co; 1981. p. 331-42.
20. Buckingham RA. The Classical Equation of State of Gaseous Helium, Neon and Argon. Proceedinges of the Royal Society of London A. 1938; 168:264–83.
21. Halgren TA. Representation of van der Waals (vdW) Interactions in Molecular Mechanics Force Fields: Potential Form, Combination Rules, and vdW Parameters. Journal of the American Chemical Society. 1992; 114:7827–43.
22. Hill TL. Steric Effects. I. Van der Waals Potential Energy Curves. Journal of Chemical Physics. 1948; 16:399–404.
23. Halgren TA. Merck Molecular Force Field. I. Basis, Form, Scope, Parameterization, and Performance of MMFF94. J Comput Chem. 1996; 17(5–6):490–519.
24. Halgren TA. Merck Molecular Force Field. II. MMFF94 van der Waals and Electrostatic Parameters for Intermolecular Interactions. J Comput Chem. 1996; 17(5–6):520–52.
25. 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 Oct; 25(13):1656–76. [PubMed: 15264259]
26. Halgren TA, Damm W. Polarizable force fields. Curr Opin Struct Biol. 2001 Apr.11:236–42. [PubMed: 11297934]
27. Ponder JW, Case DA. Force Fields for Protein Simulations. Adv Protein Chem. 2003; 66:27–85. [PubMed: 14631816]
28. Lopes PEM, Roux B, MacKerell AD Jr. 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]
29. Banks JL, Kaminski GA, Zhou R, Mainz DT, Berne BJ, Friesner RA. Parametrizing a Polarizable Force Field from Ab Initio Data. I. The Fluctuating Point Charge Model. Journal of Chemical Physics. 1999; 110:741–54.
30. Stern HA, Kaminski GA, Banks JL, Zhou R, Berne BJ, Friesner RA. Fluctuating Charge, Polarizable Dipole, and Combined Models: Parameterization from ab Initio Quantum Chemistry. Journal of Physical Chemistry B. 1999; 103:4730–7.
31. Wang ZX, 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 Apr 30.27:781–90. [PubMed: 16526038]
32. Yu HB, Geerke DP, Liu HY, van Gunsteren WE. Molecular dynamics simulations of liquid methanol and methanol-water mixtures with polarizable models. J Comput Chem. 2006; 27(13): 1494–504. [PubMed: 16838298]
33. Jorgensen WL, Jensen KP, Alexandrova AN. Polarization effects for hydrogen-bonded complexes of substituted phenols with water and chloride ion. Journal of Chemical Theory and Computation. 2007; 3(6):1987–92. [PubMed: 21132092]
34. Patel S, Davis JE, Bauer BA. Exploring Ion Permeation Enegretics in Gramicidin A Using Polarizable Charge Equilibration Force Fields. Journal of the American Chemical Society. 2009; 131(39):13890–1. [PubMed: 19788320]

35. Lopes PEM, Lamoureux G, MacKerell AD Jr. Polarizable Empirical Force Field for Nitrogen-Containing Heteroaromatic Compounds Based on the Classical Drude Oscillator. J Comput Chem. 2009; 30(12):1821–38. [PubMed: 19090564]
36. Lopes PEM, Lamoureux G, Roux B, MacKerell AD Jr. Polarizable Empirical Force Field for Aromatic Compounds Based on the Classical Drude Oscillator. J Phys Chem B. 2007 In Press.
37. Anisimov VM, Vorobyov IV, Roux B, MacKerell J, AD. Polarizable empirical force field for the primary and secondary alcohol series based on the classical Drude model. J Chem Theory Comp. 2007; 3:1927–46.
38. Anisimov VM, Vorobyov IV, Lamoureux G, Noskov S, Roux B, MacKerell AD Jr. CHARMM All-atom Polarizable Force Field Parameter Development for Nucleic Acids. Biophys J. 2004; 86:415a.
39. 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 In Press.
40. Vorobyov IV, Anisimov VM, MacKerell AD Jr. Polarizable Empirical Force Field for Alkanes Based on the Classical Drude Oscillator Model. J Phys Chem B. 2005; 109:18988–99. [PubMed: 16853445]
41. Ren P, Ponder JW. Consistent Treatment of Inter- and Intramolecular Polarization in Molecular Mechanics Calculations. J Comput Chem. 2002; 23(16):1497–506. [PubMed: 12395419]
42. Wang J, Cieplak P, Li J, Wang J, Cai Q, Hsieh M, et al. Development of Polarizable Models for Molecular Mechanical Calculations II: Induced Dipole Models Significantly Improve Accuracy of Intermolecular Interaction Energies. Journal of Physical Chemistry B. 2011; 115:3100–11.
43. Rick SW, Stuart SJ. Potentials and Algorithms for Incorporating Polarizability in Computer Simulations. Rev Comp Chem. 2002; 18:89–146.
44. Lopes, PEM.; Harder, E.; Roux, B.; MacKerell, AD, Jr. 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 Science+Business Media B.V; 2009. p. 219-57.
45. Cieplak P, Dupradeau F-Y, Duan Y, Wang J. Polarization effects in molecular mechanical force fields. Journal of Physics: Condensed Matter. 2009; 21(33):333102.
46. Tuckerman ME, Martyna GJ, Berne BJ. Molecular Dynamics Algorithm for Condensed Systems with Multiple Time Scales. Journal of Chemical Physics. 1990; 93(2):1287–91.
47. Tuckerman ME, Martyna GJ. Understanding Modern Molecular Dynamics: Techniques and Applications. J Phys Chem B. 2000; 104:159–78.
48. Jorgensen WL. Optimized Intermolecular Potential Functions for Liquid Alcohols. Journal of Physical Chemistry. 1986; 90:1276–84.
49. Durell SR, Brooks BR, Ben-Naim A. Solvent-Induced Forces Between Two Hydrophilic Groups. Journal of Physical Chemistry. 1994; 98:2198–202.
50. 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; 110:1657–66.
51. Weiner SJ, Kollman PA, Case DA, Singh UC, Ghio C, Alagona G, et al. A New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins. Journal of the American Chemical Society. 1984; 106:765–84.
52. Reiher, WE, III. PhD Thesis. Harvard University; 1985. Theoretical Studies of Hydrogen Bonding.
53. Clementi C. Coarse-grained models of protein folding: toy models or predictive tools? Current Opinion in Structural Biology. 2008; 18:10–5. [PubMed: 18160277]
54. Tozzini V. Coarse-grained models for proteins. Current Opinion in Structural Biology. 2005; 15:144–50. [PubMed: 15837171]
55. Bond PJ, Holyoake J, Ivetac A, Khalid S, Sansom MSP. Coarse-grained molecular dynamics simulations of membrane proteins and peptides. Journal of Structural Biology. 2007; 157:593–605. [PubMed: 17116404]
56. Izvekov S, Voth GA. A Multiscale Coarse-Graining Method for Biomolecular Systems. Journal of Physical Chemistry B. 2005; 109(7):2469–73.

57. DeVane R, Shinoda W, Moore PB, Klein ML. Transferable Coarse Grain Nonbonded Interaction Model for Amino Acids. Journal of Chemical Theory and Computation. 2009; 5(8):2115–24. [PubMed: 20161179]
58. Ołdziej S, Łągiewka J, Liwo A, Czaplewski C, Chinchio M, Nanias M, et al. Optimization of the UNRES Force Field by Hierarchical Design of the Potential-Energy Landscape. 3. Use of Many Proteins in Optimization. Journal of Physical Chemistry B. 2004; 108:16950–9.
59. Liwo A, KaŸmierkiewicz R, Czaplewski C, Groth M, Ołdziej S, Wawak RJ, et al. United-Residue Force Field for Off-Lattice Protein-Structure Simulations: III. Origin of Backbone Hydrogen-Bonding Cooperativity in United-Residue Potentials. J Comput Chem. 1998; 19(3):259–76.
60. Liwo A, Ołdziej S, Pincus MR, Wawak RJ, Rackovsky S, Scheraga HA. A United-Residue Force Field for Off-Lattice Protein-Structure Simulations. I. Functional Forms and Parameters of Long-Range Side-Chain Interaction Potentials from Protein Crystal Data. J Comput Chem. 1997; 18:849–76.
61. Liwo A, Pincus MR, Wawak RJ, Rackovsky S, Ołdziej S, Scheraga HA. A United-Residue Force Field for Off-Lattice Protein-Structure Simulations. II. Parameterization of Short-Range Interactions and Determination of Weights of Energy Terms by Z-Score Optimization. J Comput Chem. 1997; 18(7):874–87.
62. Liwo A, Arłukowicz P, Ołdziej S, Czaplewski C, Makowski M, Sheraga HA. Optimization of the UNRES Force Field by Hierarchical Design of the Potential-Energy Landscape. 1. Tests of the Approach Using Simple Lattice Protein Models. Journal of Physical Chemistry B. 2004; 108:16918–33.
63. Ołdziej S, Liwo A, Czaplewski C, Pillardy J, Sheraga HA. Optimization of the UNRES Force Field by Hierarchical Design of the Potential-Energy Landscape. 2. Off-Lattice Tests of the Method with Single Proteins. Journal of Physical Chemistry B. 2004; 108:16934–49.
64. Murkherjee A, Bagchi B. Contact pair dynamics during folding of two small proteins: Chicken villin head piece and the Alzheimer protein β-amyloid. Journal of Chemical Physics. 2004; 120(3): 1602–12. [PubMed: 15268287]
65. Khalili M, Saunders JA, Liwo A, Ołdziej S, Sheraga HA. A united residue force field for calcium-protein interactions. Protein Science. 2004; 13(10):2725–35. [PubMed: 15388862]
66. Smith AV, Hall CK. α-Helix Formation: Discontinuous Molecular Dynamics on an Intermediate-Resolution Protein Model. Proteins: Structure, Function and Genetics. 2001; 44(3):344–60.
67. Smith AV, Hall CK. Assembly of a tetrameric α-helix bundle: computer simulations on an intermediate-resolution protein model. Proteins: Structure, Function and Genetics. 2001; 44(3): 376–91.
68. Takada S, Luthey-Schulten Z, Wolynes PG. Folding dynamics with nonadditive forces: A simulation study of a designed helical protein and a random heteropolymer. Journal of Chemical Physics. 1999; 110(23):11616–29.
69. Takada S. Protein Folding Simulation With Solvent-Induced Force Field: Folding Pathway Ensemble of Three-Helix-Bundle Proteins. PROTEINS: Structure, Function and Genetics. 2001; 42:85–98.
70. Fujitsuka Y, Takada S, Luthey-Schulten Z, Wolynes PG. Optimizing Physical Energy Functions for Protein Folding. PROTEINS: Structure, Function and Bioinformatics. 2004; 54:88–103.
71. Bond PJ, Sansom MSP. Insertion and Assembly of Membrane Proteins via Simulation. Journal of the American Chemical Society. 2006; 128(8):2697–704. [PubMed: 16492056]
72. Wang JM, Wang W, Kollman PA, Case DA. Automatic atom type and bond type perception in molecular mechanical calculations. Journal of Molecular Graphics and Modelling. 2006; 25(2): 247–60. [PubMed: 16458552]
73. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and testing of a general AMBER force field. J Comput Chem. 2004; 25:1157–74. [PubMed: 15116359]
74. Vanommeslaeghe K, MacKerell AD Jr. Automation of the CHARMM General Force Field.
75. Vanommeslaeghe K, Raman EP, MacKerell AD Jr. Automation of the CHARMM General Force Field (CGenFF) II: assignment of bonded parameters and charges by analogy. Journal of Chemical Information and Modeling. 2012

76. Yesselman JD, Price DJ, Knight JL, Brooks CL III. MATCH: An Atom-Typing Toolset for Molecular Mechanics Force Fields. J Comput Chem. 2012; 33(2):189–202. [PubMed: 22042689]
77. Zoete V, Cuendet MA, Grosdidier A, Michielin O. SwissParam: A Fast Force Field Generation Tool for Small Organic Molecules. J Comput Chem. 2011; 32(11):2359–68. [PubMed: 21541964]
78. Halgren TA. Merck Molecular Force Field. V. Extension of MMFF94 Using Experimental Data, Additional Computational Data, and Empirical Rules. J Comput Chem. 1996; 17(5–6):616–41.
79. Bush BL, Bayly CI, Halgren TA. Consensus Bond-Charge Increments Fitted to Electrostatic Potential or Field of Many Compounds: Application of MMFF94 Training Set. J Comput Chem. 1999; 20:1495–516.
80. Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, et al. CHARMM General Force Field (CGenFF): A Force Field for Drug-Like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. J Comput Chem. 2010; 31(4):671–90. [PubMed: 19575467]
81. Gasteiger J, Marsili M. Iterative Partial Equalization of Orbital Electronegativity - Rapid Access to Atomic Charges. Tetrahedron. 1980; 36:3219–88.
82. Gilson MK, Gilson HS, Potter MJ. Fast assignment of accurate partial atomic charges: an electronegativity equilization method that accounts for alternate resonance forms. J Chem Inf Comp Sci. 2003; 43:1982–97.
83. Del Re G. A simple MO-LCAO method for the calculation of charge distributions in saturated organic molecules. Journal of the Chemical Society. 1958:4031–40.
84. Jakalian A, Bush BL, Jack DB, Bayly CI. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: I. Method. J Comput Chem. 2000; 21:132–46.
85. Jakalian A, Jack DB, Bayly CI. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: II. Parameterization and Validation. J Comput Chem. 2002; 23(16):1623–41. [PubMed: 12395429]
86. Storer JW, Giesen DJ, Cramer CJ, Truhlar DG. Class IV charge models: a new semiempirical approach in quantum chemistry. J Comput Aided Mol Des. 1995 Feb 9.:87–110. [PubMed: 7751872]
87. Thompson JD, Cramer CJ, Truhlar DG. Parameterization of charge model 3 for AM1, PM3, BLYP, and B3LYP. J Comput Chem. 2003 Aug.24:1291–304. [PubMed: 12827670]
88. Marenich AV, Jerome SV, Cramer CJ, Truhlar DG. Charge Model 5: An Extension of Hirshfeld Population Analysis for the Accurate Description of Molecular Interactions in Gaseous and Condensed Phases. Journal of Chemical Theory and Computation. 2012; 8(2):527–41.
89. Udier-Blagović M, Morales de Tirado P, Pearlman SA, Jorgensen WL. Accuracy of Free Energies of Hydration Using CM1 and CM3 Atomic Charges. J Comput Chem. 2004; 25(11):1322–32. [PubMed: 15185325]
90. Chirlian LE, Francl MM. Atomic Charges Derived from Electrostatic Potentials: A Detailed Study. J Comput Chem. 1987; 8:894–905.
91. Breneman CM, Wiberg KB. Determining Atom-Centered Monopoles From Molecular Electrostatic Potentials - The Need for High Sampling Density in Formamide Conformational-Analysis. J Comput Chem. 1990; 11:361–73.
92. Singh UC, Kollman PA. An Approach to Computing Electrostatic Charges for Molecules. J Comput Chem. 1984; 5:129–45.
93. 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. J Phys Chem. 1993; 97:10269–80.
94. Knight JL, Yesselman JD, Brooks CL III. Assessing the Quality of Absolute Hydration Free Energies among CHARMM-Compatible Ligand Parameterization Schemes. J Comput Chem. 2013; 34:893–903. [PubMed: 23292859]
95. Hura G, Russo D, Glaeser RM, Head-Gordon T, Krack M, Parrinello M. Water structure as a function of temperature from X-ray scattering experiments and *ab initio* molecular dynamics. Phys Chem Chem Phys. 2003; 5:1981–91.

96. Kučerka N, Nagle JF, Sachs JN, Feller SE, Pencer J, Jackson A, et al. Lipid Bilayer Structure Determined by the Simultaneous Analysis of Neutron and X-Ray Scattering Data. Biophysical Journal. 2008; 95:2356–67. [PubMed: 18502796]
97. Perlmutter JD, Sachs JN. Experimental verification of lipid bilayer structure through multi-scale modeling. Biochimica et Biophysica Acta. 2009; 1788:2284–90. [PubMed: 19616507]
98. 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. Journal of Physical Chemistry B. 2010; 114(23):7830–43.
99. Momany FA, Carruthers LM, McGuire RF, Scheraga HA. Intermolecular Potentials from Crystal Data. III. Determination of Empirical Potentials and Application to the Packing Configurations and Lattice Energies in Crystals of Hydrocarbons, Carboxylic Acids, Amines, and Amides. The Journal of Physical Chemistry. 1974; 78(16):1595–620.
100. Momany FA, Carruthers LM, Scheraga HA. Intermolecular Potentials from Crystal Data. IV. Application of Empirical Potentials to the Packing Configurations and Lattice Energies in Crystals of Amino Acids. The Journal of Physical Chemistry. 1974; 78(16):1621–30.
101. Okur A, Strockbine B, Hornak V, Simmerling C. Using PC clusters to evaluate the transferability of molecular mechanics force fields for proteins. J Comput Chem. 2003; 24:21–31. [PubMed: 12483672]
102. MacKerell AD Jr, Bashford D, Bellott M, Dunbrack RL Jr, Evanseck J, Field MJ, et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. Journal of Physical Chemistry B. 1998 Apr 30; 102(18):3586–616.
103. Lifson S, Warshel A. Consistent Force Field for Calculations of Conformations, Vibrational Spectra, and Enthalpies of Cycloalkane and *n*-Alkane Molecules. Journal of the American Chemical Society. 1968; 49:5116–29.
104. Wilson, EB.; Decius, JC.; Cross, PC. Molecular Vibrations. New York: McGraw-Hill; 1955.
105. Bose B, Zhao S, Stenutz R, Cloran F, Bondo PB, Bondo G, et al. Three-Bond C-O-C-C Spin-Coupling Constants in Carbohydrates: Development of a Karplus Relationship. Journal of the American Chemical Society. 1998; 120(43):11158.
106. Franks F, Dadok J, Ying S, Kay RL, Grigera JR. High-Field Nuclear Magnetic Resonance and Molecular Dynamics Investigations of Alditol Conformations in Aqueous and Non-Aqueous Solvents. Journal of the Chemical Society, Faraday Transactions. 1991; 87(4):579–85.
107. Karplus M. Contact Electron-Spin Coupling of Nuclear Magnetic Moments. Journal of Chemical Physics. 1959; 30(1):11–5.
108. Karplus M. Vicinal Proton Coupling in Nuclear Magnetic Resonance. Journal of the American Chemical Society. 1963; 85(18):2870–1.
109. Tvaroska I, Hricovíni M, Petráková E. An Attempt to Derive a New Karplus-Type Equation of Vicinal Proton-Carbon Coupling Constants for C-O-C-H Segments of Bonded Atoms. Carbohydrate Research. 1989; 189:359–62.
110. Guvench O, Hatcher ER, Venable RM, Pastor RW, MacKerell AD Jr. CHARMM Additive All-Atom Force Field for Glycosidic Linkages between Hexopyranoses. Journal of Chemical Theory and Computation. 2009; 5:2353–70. [PubMed: 20161005]
111. Hatcher ER, Guvench O, MacKerell AD Jr. CHARMM Additive All-Atom Force Field for Acyclic Polyalcohols, Acyclic Carbohydrates, and Inositol. Journal of Chemical Theory and Computation. 2009; 5(5):1315–27. [PubMed: 20160980]
112. Raman EP, Guvench O, MacKerell AD Jr. CHARMM Additive All-Atom Force Field for Glycosidic Linkages in Carbohydrates Involving Furanoses. Journal of Physical Chemistry B. 2010; 114(40):12981–94.
113. Yu W, He X, Vanommeslaeghe K, MacKerell AD Jr. Extension of the CHARMM General Force Field to Sulfonyl-Containing Compounds and Its Utility in Biomolecular Simulations. J Comput Chem. 2012; 33:2451–68. [PubMed: 22821581]
114. Williamson MP, Havel TF, Wüthrich K. Solution Conformation of Proteinase Inhibitor-IIA From Bull Seminal Plasma by H-1 Nuclear Magnetic-Resonance and Distance Geometry. Journal of Molecular Biology. 1985; 182(2):295–315. [PubMed: 3839023]

115. LeMaster DM, Kay LE, Brünger AT, Prestegard JH. Protein dynamics and distance determination by NOE measurements. FEBS Letters. 1988; 236(1):71–6. [PubMed: 2841170]
116. Hyberts SG, Goldberg MS, Havel TF, Wagner G. The solution structure of eglin c based on measurements of many NOEs and coupling constants and its comparison with X-ray structures. Protein Science. 1992; 1:736–51. [PubMed: 1304915]
117. Wüthrich K. NMR - This Other Method for Protein and Nucleic Acid Structure Determination. Acta Crystallographica Section D. 1995; 51(3):249–70.
118. Martins JC, Maes D, Loris R, Pepermans HAM, Wyns L, Willem R, et al. ^{1}H NMR Study of the Solution Structure of Ac-AMP2, a Sugar Binding Antimicrobial Protein Isolated from *Amaranthus caudatus*. Journal of Molecular Biology. 1996; 258:322–33. [PubMed: 8627629]
119. Felli IC, Brutscher B. Recent Advances in Solution NMR: Fast Methods and Heteronuclear Direct Detection. ChemPhysChem. 2009; 10(9–10):1356–68. [PubMed: 19462391]
120. Mallajosyula SS, MacKerell AD Jr. Influence of Solvent and Intramolecular Hydrogen Bonding on the Conformational Properties of O-Linked Glycopeptides. The Journal of Physical Chemistry B. 2011; 115:11215–29. [PubMed: 21823626]
121. Guvench O, Mallajosyula SS, Raman EP, Hatcher E, Vanommeslaeghe K, Foster TJ, et al. CHARMM Additive All-Atom Force Field for Carbohydrate Derivatives and Its Utility in Polysaccharide and Carbohydrate-Protein Modeling. Journal of Chemical Theory and Computation. 2011; 7(10):3162–80. [PubMed: 22125473]
122. Buck M, Bouguet-Bonnet S, Pastor RW, MacKerell AD Jr. The structural and dynamical accuracy of molecular dynamics trajectories is improved by the CMAP correction to the CHARMM22 Protein Force Field. Comparison with NMR Relaxation data for Hen lysozyme. Biophysical Journal. 2006; 90(4):L36–L8. [PubMed: 16361340]
123. Lamoureux G, MacKerell AD Jr, Roux B. A simple polarizable model of water based on classical Drude oscillators. Journal of Chemical Physics. 2003; 119:5185–97.
124. Neumann N, Steinhauser O. Computer simulation and the dielectric constant of polarizable polar systems. Chemical Physics Letters. 1984; 106(6):563–9.
125. Allen, MP.; Tildesley, DJ. Computer Simulation of Liquids. Oxford: Clarendon Press; 1987.
126. Baker CM, MacKerell AD Jr. Polarizability rescaling and atom-based Thole scaling in the CHARMM Drude polarizable force field for ethers. Journal of Molecular Modeling. 2010; 16:567–76. [PubMed: 19705172]
127. Harder E, Anisimov VM, Whitman T, MacKerell AD Jr, Roux B. Understanding the Dielectric Properties of Liquid Amides from a Polarizable Force Field. Journal of Physical Chemistry B. 2008; 112(11):3509–21.
128. MacKerell AD Jr, Karplus M. Importance of Attractive van der Waals Contributions in Empirical Energy Function Models for the Heat of Vaporization of Polar Liquids. Journal of Physical Chemistry. 1991; 95(26):10559–60.
129. MacKerell, AD., Jr; Brooks, B.; Brooks, CL., III; Nilsson, L.; Roux, B.; Won, Y., et al. CHARMM: The Energy Function and Its Paramerization with an Overview of the Program. In: Schleyer, PvR; Allinger, NL.; Clark, T.; Gasteiger, J.; Kollman, PA.; Schaefer, HF., III, et al., editors. Encyclopedia of Computational Chemistry. Chichester: John Wiley & Sons; 1998. p. 271-7.
130. Foloppe N, MacKerell AD 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.
131. Klauda JB, Brooks BR, MacKerell AD Jr, Venable RM, Pastor RW. An Ab Initio Study on the Torsional Surface of Alkanes and its Effect on Molecular Simulations of Alkanes and a DPPC Bilayer. J Phys Chem B. 2005; 109:5300–11. [PubMed: 16863197]
132. Guvench O, MacKerell AD Jr. Automated conformational energy fitting for force-field development. J Mol Mod. 2008; 14(8):667–79.
133. Yin D, MacKerell AD Jr. Combined Ab Initio/Empirical Approach for the Optimization of Lennard-Jones Parameters. J Comput Chem. 1998; 19:334–48.
134. Kaminski G, Jorgensen WL. Performance of the AMBER94, MMFF94, and OPLS-AA Force Fields for Modeling Organic Liquids. Journal of Physical Chemistry. 1996; 100:18010–3.

135. Jorgensen WL, Madura JD, Swenson CJ. Optimized Intermolecular Potential Functions for Liquid Hydrocarbons. Journal of the American Chemical Society. 1984; 106:6638–46.

136. Chen I-J, Yin D, MacKerell AD Jr. Combined Ab Initio/Empirical Optimization of Lennard-Jones Parameters for Polar Neutral Compounds. J Comput Chem. 2002; 23:199–213. [PubMed: 11924734]

137. Shirts MR, Pitera JW, Swope WC, Pande VS. Extremely precise free energy calculations of amino acid side chain analogs: Comparison of common molecular mechanics force fields for proteins. Journal of Chemical Physics. 2003; 119:5740–61.

138. Deng Y, Roux B. Hydration of Amino Acid Side Chains: Non-Polar and Electrostatic Contributions Calculated from Staged Molecular Dynamics Free Energy Simulations with Explicit Water Molecules. Journal of Physical Chemistry B. 2004; 108(42):16567–76.

139. Baker CM, Lopes PEM, Zhu X, Roux B, MacKerell AD Jr. 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):1181–98. [PubMed: 20401166]

140. Shivakumar D, Harder E, Damm W, Friesner RA, Sherman W. Improving the Prediction of Absolute Solvation Free Energies Using the Next Generation OPLS Force Field. Journal of Chemical Theory and Computation. 2012; 8:2553–8.

141. Halgren TA. MMFF VI. MMFF94s Option for Energy Minimization Studies. J Comput Chem. 1999; 20:720–9.

142. Rappé AK, Casewit CJ, Colwell KS, Goddard WA III, Skiff WM. UFF, a Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations. Journal of the American Chemical Society. 1992; 114:10024–35.

143. Matsuoka O, Clementi E, Yoshimine M. CI Study of the Water Dimer Potential Surface. Journal of Chemical Physics. 1976; 64(4):1351–61.

144. Owicki JC, Scheraga HA. Monte Carlo Calculations in the Isothermal-Isobaric Ensemble. 1. Liquid Water. Journal of the American Chemical Society. 1977; 99(23):7403–12.

145. Clark M, Cramer R, van Opdenbosh N. Validation of the General Purpose Tripos 5.2 Force Field. J Comput Chem. 1989; 10:982–1012.

146. Dauber-Osguthorpe P, Roberts VA, Osguthorpe DJ, Wolff J, Genest M, Hagler AT. Structure and Energetics of Ligand-Binding to Proteins - Escherichia Coli Dihydrofolate Reductase Trimethoprim, a Drug-Receptor System. Proteins: Structure, Function and Genetics. 1988; 4(1): 31–47.

147. Momany FA, Rone R. Validation of the General Purpose QUANTA 3.2/CHARMm Force Field. J Comput Chem. 1992; 13(7):888–900.

148. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM Jr, Ferguson DM, et al. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. Journal of the American Chemical Society. 1995; 117(19):5179–97.

149. Jorgensen WL, Maxwell DS, Tirado-Rives J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. Journal of the American Chemical Society. 1996 Nov 13; 118(45):11225–36.

150. Guvench, O.; MacKerell, AD, Jr. Comparison of Protein Force Fields for Molecular Dynamics Simulations. In: Kukol, A., editor. Molecular Modeling of Proteins. New Jersey: Humana Press, Inc; 2008. p. 63-88.

151. Levitt M. Molecular dynamics of native protein: I. Computer simulation of trajectories. Journal of Molecular Biology. 1983; 168(3):595–620. [PubMed: 6193280]

152. Levitt M, Hirshberg M, Sharon R, Daggett V. Potential energy function and parameters for simulations of the molecular dynamics of proteins and nucleic acids in solution. Computer Physics Communications. 1995 Sep; 91(1–3):215–31.

153. Daggett V. Protein folding-simulation. Chemical Reviews. 2006 May; 106(5):1898–916. [PubMed: 16683760]

154. Momany FA, Carruthers LM, McGuire RF, Scheraga HA. Intermolecular potentials from crystal data. III. Determination of empirical potentials and application to the packing configurations and

lattice energies in crystals of hydrocarbons, carboxylic acids, amines, and amides. Journal of Physical Chemistry. 1974; 78(16):1595–620.

155. Momany FA, McGuire RF, Burgess AW, Scheraga HA. Energy parameters in polypeptides. VII. Geometric parameters, partial atomic charges, nonbonded interactions, hydrogen bond interactions, and intrinsic torsional potentials for the naturally occurring amino acids. Journal of Physical Chemistry. 1975; 79(22):2361–81.
156. Némethy G, Pottle MS, Scheraga HA. Energy parameters in polypeptides. 9. Updating of geometrical parameters, nonbonded interactions, and hydrogen bond interactions for the naturally occurring amino acids. Journal of Physical Chemistry. 1983; 87(11):1883–7.
157. Némethy G, Gibson KD, Palmer KA, Yoon CN, Paterlini G, Zagari A, et al. Energy parameters in polypeptides .10. Improved geometrical parameters and nonbonded interactions for use in the ECEPP/3 algorithm, with application to proline-containing peptides. Journal Of Physical Chemistry. 1992 Jul 23; 96(15):6472–84.
158. Arnautova YA, Jagielska A, Sheraga HA. A New Force Field (ECEPP-05) for Peptides, Proteins, and Organic Molecules. Journal of Physical Chemistry B. 2006; 110:5025–44.
159. Arnautova YA, Vorobjev YN, Vila JA, Sheraga HA. Identifying native-like protein structures with scoring functions based on all-atom ECEPP force fields, implicit solvent models and structure relaxation. Proteins: Structure, Function and Bioinformatics. 2009; 77(1):38–51.
160. Oldziej, S.; Czaplewski, C.; Liwo, A.; Chinchio, M.; Nanias, M.; Vila, JA., et al. Physics-based protein-structure prediction using a hierarchical protocol based on the UNRES force field: Assessment in two blind tests. Proceedings Of The National Academy Of Sciences Of The United States Of America; 2005 May 24; p. 7547-52.
161. Arnautova YA, Sheraga HA. Use of Decoys to Optimize an All-Atom Force Field Including Hydration. Biophysical Journal. 2008; 95:2434–49. [PubMed: 18502794]
162. Case DA, Cheatham TE III, Darden T, Gohlke H, Luo R, Merz KM Jr, et al. The Amber biomolecular simulation programs. J Comput Chem. 2005; 26:1668–88. [PubMed: 16200636]
163. Kaminski G, 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; 105:6474–87.
164. Price MLP, Ostrovsky D, Jorgensen WL. Gas-Phase and Liquid-State Properties of Esters, Nitriles, and Nitro Compounds with the OPLS-AA Force Field. J Comput Chem. 2001; 22:1340–52.
165. Jorgensen WL, McDonald NA. Development of an all-atom force field for heterocycles. Properties of liquid pyridine and diazenes. Journal of Molecular Structure (Theochem). 1998; 424(1–2):145–55.
166. Mohamadi F, Richards NGJ, Guida WC, Liskamp R, Lipton M, Caufield C, et al. MacroModel - an integrated software system for modeling organic and bioorganic molecules using molecular mechanics. J Comput Chem. 1990; 11(4):440–67.

Table 1

Force field parameters in the class 1 additive potential energy function (equation 1). The Lennard-Jones parameters are further explained in subsection 3.3.

Symbol	Meaning
b_0	Reference bond length
K_b	Bond force constant
θ_0	Reference valence angle
K_θ	Angle force constant
ϕ_0	Improper dihedral angle reference value (usually 0)
K_ϕ	Improper dihedral force constant
n	Dihedral multiplicity
δ_n	Dihedral phase
$K_{\varphi n}$	Dihedral amplitude
q_i, q_j	Partial charges
$R_{\min,ij}$	Lennard-Jones radius
ε_{ij}	Lennard-Jones well depth