2 Empirical Force Fields

Alexander D. MacKerell, Jr.

Protein structure and dynamics and, therefore, their biological functions are dictated by a collection of forces that vary from those associated with covalent linkages, such as bonds, to long-range through space forces, such as electrostatic or coulombic interactions. Accordingly, to be able to apply theoretical approaches to understand the behavior of proteins, it is necessary to be able to accurately predict the change in energy of a protein as a function of the change in conformation. Importantly, such predictions must include contributions from the environment in which the protein is immersed. While quantum-mechanical (QM) methods are attractive in their ability to model complex chemical phenomena at the level of electronic structure, such methods are typically inappropriate for proteins due to the large size of these macromolecules as well as the need to treat their environment in an explicit fashion. Rather, molecular mechanics (MM), which rely on potential energy functions or empirical force fields, afford the computational speed to allow for calculations on proteins along with their environment.

2.1 Potential Energy Functions

The computational speed associated with molecular mechanics is based on the simplicity of the mathematical models used in the potential energy function to relate the structure of the system to its energy. This simplicity is based on the smallest particles in the model typically being atoms, which are treated as point masses centered on the nucleus of each atom in the molecules comprising the system under study. The potential energy function therefore describes the interactions between the atoms in the system.

An example of the potential energy function used in the additive CHARMM force fields (Brooks et al., 1983; MacKerell et al., 1998b) is shown in Eq. (2.1); similar energy functions are used in the common macromolecular force fields for proteins including OPLS/AA (Jorgensen and Tirado-Rives, 1988), AMBER (Cornell

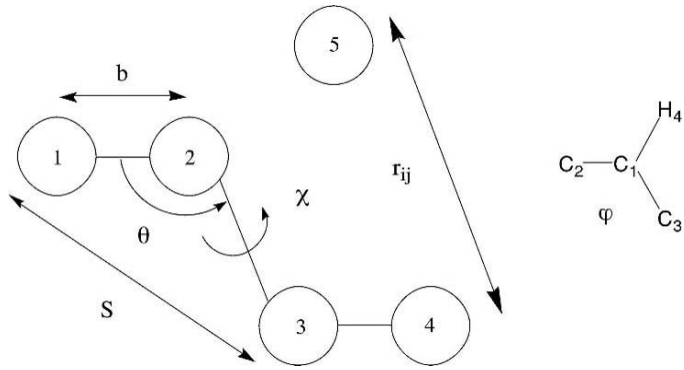

Fig. 2.1 Schematic diagram of the terms used to describe the energy as a function of the conformation in the potential energy function. The bond length between two covalently attached atoms 1 and 2 is b, θ is the valence angle between atoms 1, 2, and 3, χ is the dihedral angle involving atoms 1, 2, 3, and 4, S is the Urey-Bradley distance between atoms 1 and 3, and r_{ij} is the through space distance between atoms 1 and 4. The inset shows an example of an improper dihedral, φ, which is defined as the dihedral C1–C2–C3–H4.

et al., 1995), and GROMOS (van Gunsteren, 1987).

$$U(\vec{R}) = \sum_{\text{bounds}} K_b(b - b_0)^2 + \sum_{\text{angle}} K_\theta(\theta - \theta_0)^2 + \sum_{\text{UB}} K_{\text{UB}}(S - S_0)^2 + \sum_{\text{dihedrals}} K_\chi(1 + \cos(n\chi - \delta)) + \sum_{\text{impropers}} K_{\text{imp}}(\varphi - \varphi_0)^2 + \sum_{nonbound} \varepsilon_{ij}\left[\left(\frac{R_{\min,ij}}{r_{ij}}\right)^{12} - 2\left(\frac{R_{\min,ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{4\pi e r_{ij}} \quad (2.1)$$

Equation (2.1), where the potential energy, U, is calculated as a function of the atomic positions, $\vec{R}$, includes terms for the internal (i.e., bonded) and external (i.e., interaction or nonbond) contributions. Internal terms include the bonds, valence angles, Urey-Bradley, dihedral or torsion angles, and improper dihedral terms while the external terms include the van der Waals (vdW) interations, treated via the Lennard-Jones (LJ) 6–12 term and the electrostatic interactions. In Eq. (2.1), terms describing the geometry of the molecule include the bond length, b, the valence angle, θ, the distance between atoms separated by two covalent bonds (Urey-Bradley term, 1,3 distance), S, the dihedral or torsion angle, χ, the improper angle, φ, and the distance between atoms i and j, r_{ij}. The schematic diagram in Fig. 2.1 illustrates the terms included in Eq. (2.1).

In order for the potential energy function to represent different types of, for example, bonds (e.g., C–C single versus double bonds) or atom types, parameters are used for each type of bond, angle, atom type, and so on in the molecules in the system. These parameters include the bond force constant and equilibrium distance, K_b and

b_0, respectively; the valence angle force constant and equilibrium angle, K_θ, and θ_0; the Urey-Bradley force constant and equilibrium distance, K_{UB} and S; the dihedral angle force constant, multiplicity, and phase angle, K_χ, n, and δ; and the improper force constant and equilibrium improper angle, K_ϕ and ϕ_0. External parameters that describe the interactions between atoms i and j include the partial atomic charges, q_i, and the LJ well-depth, ε_{ij}, and minimum interaction radius, $R_{\min,ij}$, used to treat the vdW interactions. Typically, ε_i and $R_{\min,i}$ are obtained for individual atom types and then combined to yield ε_{ij} and $R_{\min,ij}$ for the interacting atoms via combining rules. In CHARMM, ε_{ij} values are obtained via the geometric mean ($\varepsilon_{ij} = \text{sqrt}(\varepsilon_i * \varepsilon_j)$) and $R_{\min,ij}$ via the arithmetic mean, $R_{\min,ij} = (R_{\min,i} + R_{\min,j})/2$. The dielectric constant, e, is set to one in all calculations where solvent is considered explicitly (see below), corresponding to the permittivity of vacuum.

Essential for the modeling of proteins, as well as all biomolecules, is the proper treatment of hydrogen bonding. Earlier force fields included explicit terms for hydrogen bonds (Weiner and Kollman, 1981); however, it has been shown that the combination of the Lennard-Jones and coulombic interactions produces an accurate representation of both the distance and angle dependencies of hydrogen bonds (Reiher, 1985). This success has allowed for the omission of explicit terms to treat hydrogen bonding from the majority of empirical force fields. It should be noted that the LJ and electrostatic parameters are highly correlated, such that LJ parameters determined for a set of partial atomic charges will not be applicable to another set of charges. Moreover, the internal parameters are dependent on the external parameters. For example, the barrier to rotation about the C–O bond in ethanol includes contributions from the electrostatic and vdW interactions between the hydroxyl hydrogen and the rest of the molecule as well as contributions from the bond, angle, and dihedral terms. Thus, if the LJ parameters or charges are changed, the internal parameters will have to be reoptimized to produce the correct energy barrier. Finally, condensed phase properties obtained from empirical force field calculations contain contributions for the conformations of the molecules being studied as well as external interactions between those molecules, emphasizing the importance of accurate treatment of both internal and external portions of the force field for accurate condensed phase simulations.

Beyond the terms included in Eq. (2.1), additional terms may be included in a potential energy function; such extended energy functions are typically referred to as Class II force fields. Class II force fields can include higher order corrections for the bond and valence angle terms and/or cross terms between, for example, bonds and valence angles or valence angles and dihedrals (Lii and Allinger, 1991; Derreumaux and Vergoten, 1995; Halgren, 1996a; Sun, 1998; Ewig et al., 2001; Palmo et al., 2003). Other alternative terms include the use of a Morse function for bonds (Burkert and Allinger, 1982). This function allows for bond breaking to be included in an empirical force field. Another alternative is the use of a cosine-based valence angle term that is well behaved for near-linear valence angles (Mayo et al., 1990; Rappé et al., 1992). For the dihedral term a recent improvement that avoids singularities associated with derivatives of torsion angle cosines and allows for application of

any value of the phase has been presented (Blondel and Karplus, 1996) and, more recently, the introduction of a two-dimensional (2D) grid-based dihedral energy correction map (CMAP) (MacKerell et al., 2004a,b) that allows for any 2D dihedral surface (e.g., a QM ϕ, ψ surface of the alanine dipeptide) to be reproduced nearly exactly by the force field (see below). These two terms are included in the recent version of the CHARMM force field for proteins. Typically inclusion of these terms in an energy function increases its accuracy in treating conformational energies, especially at geometries far from the minimum-energy or equilibrium values as well as yield improved treatment of vibrational spectra. However, it should be emphasized that Class I force fields [i.e., those based on Eq. (2.1)] can yield accuracies similar to the Class II force fields when the parameters are properly optimized. In general, Class I force fields, when applied to biomolecular simulations in the vicinity of room temperature, adequately treat the intramolecular distortions, including relative conformational energies associated with large structural changes.

The external portion of a potential energy function may also be extended beyond that in Eq. (2.1), including alternate forms of both the vdW interactions and the electrostatics. The three primary alternatives to the LJ 6–12 term included in Eq. (2.1) are designed to "soften" the repulsive wall associated with Pauli exclusion. For example, the Buckingham potential (Buckingham and Fowler, 1985) uses an exponential term to treat repulsion while a buffered 14–7 term is used in the MMFF force field (Halgren, 1996b). A simple alternative is to replace the r^{12} repulsion with an r^9 term. All of these forms more accurately treat the repulsive wall as judged by high-level QM calculations (Halgren, 1992). However, as with the harmonic internal terms in Class I force fields, the LJ term appears to be adequate for biomolecular simulations at or near room temperature.

2.2 Implementation of Potential Energy Functions

As stated above, once an empirical force field is available, it may be used, in combination with the necessary software, to calculate the change of energy of a system as a function of coordinates. However, more useful is the combination of an empirical force field with numerical approaches allowing for sampling of relevant conformations via, e.g., a molecular dynamics (MD) simulation to be performed (Tuckerman and Martyna, 2000). Such approaches can be used to predict a variety of structural and thermodynamic properties, including free energies, via statistical mechanics (McQuarrie, 1976). Importantly, such approaches allow for comparisons with experimental thermodynamic data and the atomic details of interactions between molecules that dictate the thermodynamic properties can be obtained. Such atomic details are often difficult to access via experimental approaches, motivating the application of computational approaches.

Proper application of an empirical force field when performing MD simulations or other calculations on proteins is an essential consideration. Due to the central role of the external interactions in the energy function, it is important that all nonbond

interactions between all atom–atom pairs be considered. The Ewald method can be used to treat the long-range electrostatic interactions for periodic systems (Ewald, 1921). Recent variations of the Ewald method that are computationally more tractable include the particle Mesh Ewald approach (Darden, 2001). Alternatively, reaction field methods can be used to simulate finite (e.g., spherical) systems (Beglov and Roux, 1994; Bishop et al., 1997; Im et al., 2001). Concerning the vdW or LJ interactions, the long-range contributions to this term beyond the atom–atom truncation distance (i.e., those beyond a distance where the atom–atom interactions are calculated explicitly) can be corrected for by assuming those contributions are homogeneous in nature (Allen and Tildesley, 1989; Lague et al., 2004).

The integrators that generate proper ensembles in MD simulations are another important consideration, as attaining the proper ensemble in an MD simulation is essential for direct comparison with experimental data (Tuckerman et al., 1992; Martyna et al., 1994; Feller et al., 1995; Barth and Schlick, 1998; Tuckerman and Martyna, 2000). Extensions of MD simulations have been developed that significantly increase the sampling of conformational space including locally enhanced sampling (Elber and Karplus, 1990; Hansmann, 1997; Simmerling et al., 1998) and replica-exchange or parallel tempering (Hansmann, 1997; Sugita and Okamoto, 1999; Nymeyer et al., 2004). It should be noted that the deterministic nature of MD simulations is typically lost when using such approaches, although the replica-exchange method can produce results that correspond to a proper ensemble. As always, the appropriate use of these different methods greatly facilitates investigations of molecular interactions via condensed phase simulations.

2.3 Treatment of Aqueous Solvation

Protein structure and function is greatly influenced by the condensed phase (i.e., aqueous) environment in which they exist. Accordingly, an empirical force field for proteins must treat the condensed phase environment in an accurate manner. Treatment of the protein environment may be performed using either explicit or implicit models. Explicit models, where the water, ions, and so on, are included explicitly in the simulation, are more microscopically accurate while implicit or continuum models can produce savings in computer time over explicit models and have the advantage of directly yielding free energies of solvation.

A number of explicit water models have been used in protein simulations including the TIP3P, TIP4P (Jorgensen et al., 1983), SPC, extended SPC/E (Berendsen et al., 1987) and F3C (Levitt et al., 1997) models. TIP3P is the most commonly used water model. Its three-point design (i.e., one oxygen and two hydrogen atoms) makes it computationally tractable and it yields the correct thermodynamic properties of water. Structurally the model treats the first and second solvation shells with reasonable accuracy. However, the second or tetrahedral peak in the O–O radial distribution is underestimated and the diffusion constant of the model is significantly larger than the corresponding experimental value (Feller et al., 1996). Another widely used

three-point model is SPC. This model uses an internal tetrahedral geometry (i.e., H–O–H angle $= 109.47°$) leading to increased structure over TIP3P, as evidenced by a more-defined tetrahedral peak in the O–O radial distribution function. A variant of SPC, the SPC/E model, includes a correction for the polarization self-energy that yields improved structure and diffusion properties. However, this correction leads to an overestimation of the water potential energy in the bulk phase, which may perturb the energetic balance of solvent–solvent, solute–solvent, and solute–solute interactions. This problem must be considered when using this model in biomolecular simulations. TIP4P is a four-point water model that includes an additional particle along the H–O–H bisector. The additional particle overcomes many of the limitations listed above, although the computer demands of the model are higher. Recently, new water models have been presented (Mahoney and Jorgensen, 2000; Glättli et al., 2003), although they have not seen wide use in biomolecular simulations.

Selection of the proper water model is important for a successful simulation. The most important consideration is the compatibility of the model with the biomolecular force field being used. Such compatibility is important due to most force fields being developed in conjunction with a specific water model (e.g., AMBER, OPLS and CHARMM with TIP3P, OPLS also with TIP4P, GROMOS with SPC, ENCAD with F3C), such that it is best to use a force field with its prescribed water model unless special solvent requirements are important.

Implicit solvation models treat the protein environment as a continuum, for example, by treating regions not "inside" the protein with the dielectric constant of water (Davis and McCammon, 1990; Honig, 1993). Such models offer significant computational savings while yielding reasonably accurate treatment of solvation. Accordingly, implicit models are useful when extensive sampling of conformational space is required, as in protein folding. However, these models can fail when highly specific water–biomolecule interactions have an important structural or energetic role. The most widely used implicit solvation models are Poisson-Boltzmann (PB) and generalized Born (GB) models. In the PB model, contributions from solvent polarization along with the asymmetric shapes of biological molecules are taken into account (Gilson and Honig, 1988), from which free energies of solvation may be determined. Advances in this approach have included the optimization of atomic radii to reproduce experimental free energies of solvation of model compounds representative of proteins (Nina et al., 1997; Banavali and Roux, 2002). GB approaches (Still et al., 1990) are an alternative to PB that also yield free energies of solvation while being less computationally expensive, thereby facilitating their use in MD simulations. Several GB models have been developed that yield free energies of solvation at a level of accuracy similar to PB methods (Schaefer and Karplus, 1996; Jayaram et al., 1998; Onufriev et al., 2000; Zhang et al., 2001; Lee et al., 2003). Both the PB and GB methods can be combined with free energy solvent accessibility (SA) terms that account for the hydrophobic effect (Qui et al., 1997; Gallicchio et al., 2003), referred to as PB/SA or GB/SA approaches. Recent developments based on the GB method involve an improved treatment of vdW dispersion contributions beyond the typical solvent accessibility related terms (Gallicchio and Levy, 2004). Other implicit

models that have been used in biomolecular simulations include the Langevin Dipoles Model (Florián and Warshel, 1997) and the EEF1 model (Lazaridis and Karplus, 1999). More information on implicit solvating models can be obtained from a recent review by Feig and Brooks (Feig and Brooks, 2004).

The PB/SA and GB/SA methods can be used for postprocessing of trajectories from MD simulations to obtain free energies of solvation. In this approach an MD simulation of the biomolecule(s) is performed using an explicit solvent representation followed by estimation of the free energy of solvation using the solute coordinates from the simulation (i.e., biomolecule only with the solvent omitted) (Kollman et al., 2000). This allows for determination of the free energy of solvation of a biomolecule averaged over the length of a simulation, using structures obtained with an explicit solvent representation. This approach is particularly attractive for the calculation of free energies of binding of macromolecular complexes (Jayaram et al., 2002; Gohlke et al., 2003; Habtemariam et al., 2005). This type of approach also has great utility for the estimation of ligand–protein binding (Ferrara et al., 2004), at a computationally reasonable cost as required for testing of large numbers of drug candidates.

2.4 Empirical Force Field Optimization

The ability of a simple potential energy function such as that in Eq. (2.1) to accurately model the energies as a function of protein conformation is based on proper optimization of the parameters used in the energy function. Indeed, until the parameters are available, one does not truly have an empirical force field. And the quality of that empirical force field is judged by its ability to accurately reproduce the experimental regimen.

Parameter optimization is based on reproducing a set of target data, including information on small model compounds representative of proteins as well as on proteins themselves. Target data are ideally obtained from experiments, though a majority of the data are often obtained from QM calculations. QM calculations are readily applicable to most small molecules; however, limitations in QM level of theory, especially with respect to the treatment of dispersion interactions (Chalasinski and Szczesniak, 1994; Chen et al., 2002), require the use of experimental data when available (MacKerell, 2004).

Details on the optimization of internal parameters have been presented previously by a number or workers (Halgren, 1996c; Ewig et al., 2001; MacKerell, 2001; Wang and Kollman, 2001). Briefly, the equilibrium bond, valence angle, and Urey-Bradley parameters along with the dihedral multiplicity and phase are optimized to reproduce internal geometries of the model compounds. The target data are often QM data, although it has been shown that condensed phase effects can influence the internal geometry of a molecule, such that survey data from structures in the Cambridge Structural Database (CSD, http://www.ccdc.cam.ac.uk/) (Allen et al., 1979) may be considered the ideal. The value of such data for treatment of

the peptide bond has previously been discussed (MacKerell et al., 1998; MacKerell, 2004). Force constants for the bond, valence angle, Urey-Bradley, dihedral angle, and improper angles are optimized to reproduce vibrational spectra, including both the frequencies as well as the potential energy distribution (PED) (i.e., the contribution of internal degrees of freedom to the individual frequencies). Again the ideal data are obtained from condensed phase vibrational studies, although such data are typically limited making vibrational data from QM calculations the most commonly used. It should be emphasized that QM vibrational analysis allows for detailed assignment of the PED and, even when good experimental data are available, QM calculations are often advantageous to perform the assignments. When performing optimization of vibrational spectra, it should be noted that the low-frequency modes represent the largest structural distortions that occur in a molecule, such that their proper treatment is important for accurately treating the structural distortions that occur during MD simulations. Conformational energies from QM calculations, including barrier heights for rotations about dihedrals, are typically used for the final optimization of the dihedral angle parameters. In the CHARMM force fields the dihedral parameters are initially optimized based on vibrational data with only the parameters associated with dihedrals that involve all nonhydrogen atoms adjusted to reproduce potential energy surfaces. This final optimization is again important as the rotations about dihedrals represent the largest structural changes that occur in MD simulations of proteins. Recent work on lipids emphasizes the importance of proper treatment of the conformational energies (Klauda et al., 2005). In addition, empirical optimization of dihedral parameters to reproduce experimental distributions of conformers, such at the phi, psi angle distributions in proteins, have been shown to be important (MacKerell et al. 2004a,b). Those efforts have included optimization of the grid-based energy correction map discussed below.

Significant effort by a number of groups has gone into the determination of the electrostatic parameters; the partial atomic charges, q_i. Of the methods currently in use, the most common methods for proteins are the supramolecular and QM electrostatic potential (ESP) approaches. Other variations include bond charge increments (Bush et al., 1999; Jakalian et al., 2000) and electronegativity equilization methods (Gilson et al., 2003), although these methods are typically applied to small, drug-like molecules. An important consideration with the determination of partial atomic charges, related to the Coulombic treatment of electrostatics, is the omission of the explicit treatment of electronic polarizability. Due to this omission, it is necessary for static, partial atomic charges to reproduce the average polarization that occurs in the condensed phase environment. This is achieved by "enhancing" the charges of a molecule leading to an overestimation of the dipole moment as compared to the gas phase value. This is referred to as an implicitly polarized model. For example, many of the water models used in protein empirical force fields (e.g., TIP3P, TIP4P, SPC) have dipole moments in the vicinity of 2.2 debye (Jorgensen et al., 1983), versus the gas phase value of 1.85 debye. Inclusion of implicit polarizability allows for empirical force fields based on Eq. (2.1), which are often referred to as additive, to reproduce a variety of condensed phase properties (Rizzo and Jorgensen, 1999).

These additive models have been extensively used for simulations of proteins, as well as other biological molecules; however, they are limited in that they do not reproduce the change in electrostatic interactions due to inductive effects associated with changes in the polarity of the environment.

The supramolecular approach for the determination of partial atomic charges is used in the OPLS (Jorgensen and Tirado-Rives, 1988; Jorgensen et al., 1996) and CHARMM (MacKerell et al., 1998b; Foloppe and MacKerell, 2000; Feller et al., 2002) force fields. This approach involves optimization of the charges to reproduce QM-determined minimum interaction energies and geometries of a model compound with individual water molecules or for model compound dimers. Typically, the HF/6-31G* level of theory was used for the QM calculations, due to its overestimation of dipole moments (Cieplak et al., 1995), leading to the implicitly polarizable model discussed above. An additional advantage of the supramolecular approach is that in the QM calculation, local polarization effects due to the charge induction caused by the two interacting molecules are included, facilitating determination of charge distributions appropriate for the condensed phase.

It should be noted that although it has recently been shown that QM methods can accurately reproduce gas phase experimental interaction energies for a range of model compound dimers (Kim and Friesner, 1997; Huang and MacKerell, 2002), it is important to maintain the QM level of theory that was historically used for a particular force field when extending that force field to novel molecules. This assures that the balance of the nonbond interactions between different molecules in the system being studied is maintained. Finally, when considering the transferability of charges obtained from the supramolecular approach, it should be noted that the charges are typically obtained for functional groups such that they may be directly transferred between molecules.

The other commonly applied approach for charge determination in empirical force fields is ESP charge fitting. This methodology is based on the adjustment of charges to reproduce a QM-determined ESP mapped onto a grid surrounding the model compound. ESP methods are widely used and a number of charge fitting methods based on this approach have been developed (Singh and Kollman, 1984; Chirlian and Francl, 1987; Merz, 1992; Bayly et al., 1993; Henchman and Essex, 1999). Application of ESP fitting approaches is hampered by difficulties in unambiguously fitting charges to an ESP (Francl et al., 1996) and charges on "buried" atoms (e.g., a carbon to which three or four nonhydrogen atoms are covalently bound) tend to be underdetermined, requiring the use of restraints during fitting (Bayly et al., 1993). The latter method is referred to as Restrained ESP (RESP). In addition, the QM ESP is typically determined via gas phase calculations, which may yield charges that are not consistent with the condensed phase. Recent developments are addressing this limitation (Laio et al., 2002). Another problem is that multiple conformations of flexible molecules must also be taken into account (Cieplak et al., 1995), although it should be noted that the last two problems are also present to varying extents in the supramolecular approach. For ESP fitting, the QM level of theory has historically been HF/6-31G*, as used in the AMBER force fields (Cornell et al., 1995), although

higher level QM calculations have been applied more recently in conjunction with the RESP approach (Duan et al., 2003). In summary, the supramolecular and ESP methods are both useful for the determination of partial atomic charges and, as with the water models, the method to use should be that which is consistent with the remainder of the force field.

The most difficult aspect of empirical force fields to optimize are the LJ terms, although proper treatment of these terms is essential for obtaining accurate condensed phase properties from empirical force fields. A big part of the difficulty in optimizing LJ parameters are limitations in the quality of QM calculations in treating dispersion interactions (Chalasinski and Szczesniak, 1994; Chen et al., 2002), requiring the use of experimental condensed phase data as the target data (Jorgensen, 1984, 1986). LJ parameters are generally optimized to reproduce experimentally measured values such as heats of vaporization, densities, isocompressibilities, and heat capacities of small model compound pure solvents. Alternative target data include heats or free energies of aqueous solvation, partial molar volumes or heats of sublimation and lattice geometries of crystals (Warshel and Lifson, 1970; MacKerell et al., 1995). While these approaches have acted as the basis for several protein force fields, it should be emphasized that LJ parameters optimized in this fashion are underdetermined due to the small number of experimental observables available for the optimization of a significantly larger number of LJ parameters. This leads to the parameter correlation problem where LJ parameters for different atoms in a molecule (e.g., H and C in ethane) can compensate for each other (MacKerell, 2001). The parameter correlation problem with respect to LJ parameters has been addressed via an approach that determines the absolute values of the LJ parameters based on experimental data, as above, while their relative values are optimized using high-level QM data as the target data (Yin and MacKerell, 1996; Chen et al., 2002). In general, determination of LJ parameters is quite time consuming; however, in many instances it is feasible to directly transfer the LJ parameters between functional groups in the context of different molecules.

2.5 Protein Force Fields

Current protein MD simulations are typically performed using additive all-atom protein models, including the OPLS/AA (Jorgensen and Tirado-Rives, 1988), CHARMM22 (MacKerell et al., 1998b), and AMBER (PARM99) (Cornell et al., 1995) force fields. An alternative is the use of extended or united atom models, where the nonpolar hydrogens are treated as part of the carbon to which they are covalently bound, with polar hydrogens important for hydrogen bonding included. United atom models offer computational savings over all-atom models, and are often used with implicit solvent models. Details of the approaches used for development of the commonly used all-atom models—CHARMM22, AMBER, and OPLS—are summarized in the following paragraphs, with a brief discussion of the extended atom models given below.

Internal parameters for AMBER and CHARMM22 were derived via reproduction of both experimental and QM data for small model compounds, including reproduction of geometries and vibrational spectra. The internal parameters for the OPLS force field (Jorgensen and Tirado-Rives, 1988) were initially taken from AMBER and have been subsequently optimized to reproduce conformational energies from QM calculations yielding OPLS/AA (Jorgensen et al., 1996). Additional optimization of selected torsions has been performed using higher level QM target data (Kaminski et al., 2001). Supporting the quality of these force fields in the treatment of proteins are MD simulations, showing the three force fields to reproduce selected experimental structures in a similar fashion (Price and Brooks, 2002).

Care was taken in the optimization of the external aspects of the three force fields. In OPLS/AA and CHARMM22, the partial atomic charges are based on HF/6-31G* supramolecular data while the standard AMBER release (PARM99) is based on RESP charges fit to the same level of theory. Condensed phase simulations were used as target data for the optimization of the LJ parameters in all three force fields. In CHARMM22 and AMBER, the charges were optimized to be consistent with the TIP3P water model, whereas OPLS was developed to work with the TIP3P, TIP4P, and SPC models. Based on water dimer interaction energies, it may be anticipated that the TIP4P and SPC models will also work well with CHARMM22 and AMBER, although rigorous tests have yet to be performed. Despite similarities in the optimization of the charges for the three force fields, differences in the local charge distributions have been noted (Ponder and Case, 2003). Such differences in charges may lead to differences in the balance of the local interactions (e.g., relative hydrogen bonding strength at the peptide bond NH versus CO), which may impact the atomic details of interactions obtained from the three force fields. Based on the presence of such differences, results from MD simulations using these force fields should be interpreted taking into account the applied optimization approach.

In all of the protein force fields, a significant and ongoing effort has been made with respect to the treatment of the Ramachandran map or ϕ, ψ energy surface (Ramachandran, et al., 1963). This is due to the conformational energy as a function of the ϕ, ψ dihedral angles dictating the region of conformational space being sampled in peptide and protein simulations. The quintessential model compound for optimizing the dihedral parameters related to ϕ, ψ is the alanine dipeptide (Fig. 2.2), along with related compounds such as the glycine dipeptide and the proline dipep-

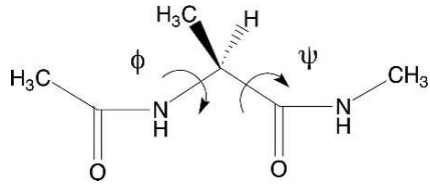

Fig. 2.2 Diagram of the alanine dipeptide including the ϕ, ψ dihedral angles used to define the Ramachandran diagram.

tide. The size of these compounds is such that they are accessible to high-level QM calculations (Head-Gordon et al., 1991; Beachy et al., 1997; Ono et al., 2002; Vargas et al., 2002; Duan et al., 2003; MacKerell et al., 2004b), the data from which can be used as target data for the parameter optimization. While targeting QM energetic data for the optimization is simple and well defined, studies have shown that directly reproducing gas phase QM data can lead to systematic problems in the conformational properties of the protein backbone (MacKerell et al., 1998b). This was shown in great detail using the CMAP approach for the treatment of the conformational energies of ϕ, ψ (MacKerell et al., 2004a,b). To overcome this limitation with the use of QM data, it has been shown that empirical adjustment of the dihedral or CMAP terms to better reproduce ϕ, ψ distributions from surveys of the protein databank (PDB) (Berman et al., 2002) leads to significant improvement in the treatment of the conformational properties of the backbone. Indeed, the recently published CHARMM22/CMAP all-atom protein force field represents a significant improvement over current force fields for proteins with respect to the treatment of both structural (Freedberg et al., 2004; MacKerell et al., 2004; Steinbach, 2004) and dynamic (Buck et al., 2006) properties.

To better understand the improvements in the treatment of the protein backbone conformational properties associated with the use of CMAP, Fig. 2.3 shows ϕ, ψ potentials of mean force (PMFs) and distributions for proteins (upper frames) and for the alanine dipeptide (lower frames). PMFs, or free energy surfaces, were obtained from MD simulations of proteins in their crystal environments using the CHARMM22 (MacKerell et al., 1998b) and CMAP modified CHARMM22 (MacKerell et al. 2004a,b) energy functions along with distributions from a survey of the PDB (Dunbrack and Cohen, 1997; Dunbrack, 2002). Comparison of the three surfaces shows the CMAP PMF to better reproduce the shape of the surface derived from the PDB. Notable are the improvements in the overall shape of the beta sheet ($\phi \sim -120^\circ, \psi \sim 150^\circ$) and alpha-helical ($\phi \sim -60^\circ, \psi \sim -40^\circ$) regions versus CHARMM22. Importantly, the improvements also occur at the model compound level, where the overall shape of the distribution of ϕ, ψ in MD simulations of the alanine dipeptide using CMAP is in better agreement with the distributions from a QM/MM simulation (SCCDFT) (Hu et al., 2003). The improved agreement at both the protein and model compound levels indicates that the overall force field is well balanced, such that the proper treatment of local contributions as evidenced by the alanine dipeptide MD results leads to the desired behavior at the macromolecular (i.e., protein) level by the force field.

Finally, it should be emphasized that the protein force fields are undergoing continual optimization in the context of the Class I energy function as well as via extension of the force field via, for example, the CMAP approach. Motivating such additional optimization is the ability to access additional target data by which to judge the quality of the force fields as well as improved algorithms and computational resources, allowing for more rigorous tests of the force fields. For example, the free energies of solvation of model compounds representative of protein side chains have recently been calculated for the AMBER, CHARMM, and OPLS/AA force

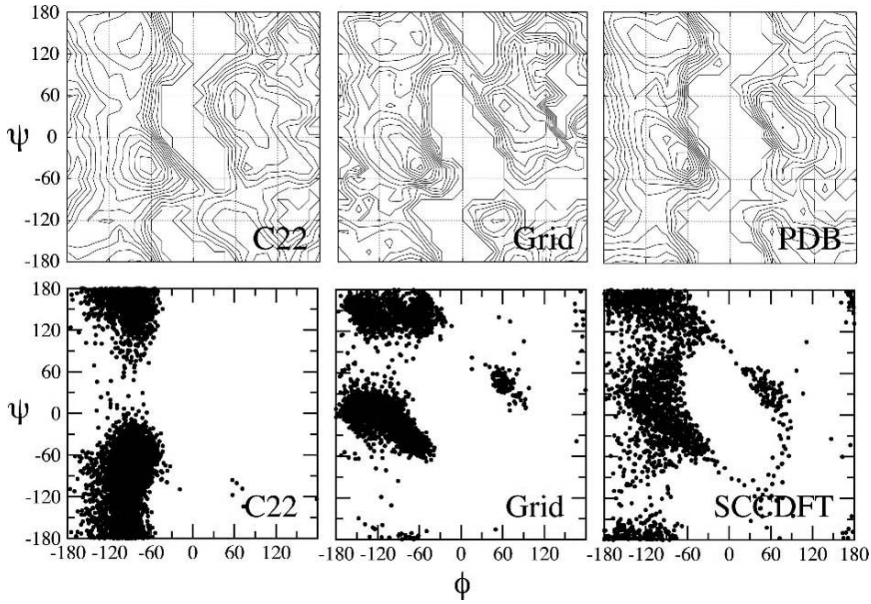

Fig. 2.3 $\phi\psi$ PMFs based on MD simulations using the CHARMM22 and CHARMM22 grid-corrected empirical force fields and from a survey of the PDB (upper frames) and $\phi\psi$ distributions from MD simulations of the alanine dipeptide (Ace-Ala-Nme, lower frames) in solution using the CHARMM22 (MacKerell et al., 1998) and CHARMM22 grid-corrected empirical force fields and previously published data from a QM/MM model (SCCDFT). PMF contours are in 0.5 kcal/mol increments up to 6 kcal/mol above the global minimum. PMFs were obtained from the respective probability distributions based on a Boltzmann distribution (McQuarrie, 1976). See reference (MacKerell et al., 2004a) for more details. Reproduced with permission from *J. Am. Chem. Soc.* 2004, 126:698–699. Copyright 2004 American Chemical Society.

fields and compared with experiment (Shirts et al., 2003); similar studies have been reported elsewhere (Villa and Mark, 2002; MacCallum and Tieleman, 2003; Deng and Roux, 2004). These studies show the force fields to yield reasonable free energies of solvation, although the need for improvements is evident.

Examples of recent adjustments of the protein force fields are numerous. Motivated by the free energy of solvation results discussed in the preceding paragraph, adjustments have been made to the tryptophan side chain parameters in CHARMM (Macias and MacKerell, 2005). Multiple adjustments have been made to the AMBER Cornell et al. force field (i.e., PARM94) (Cornell et al., 1995). In recent years, optimization of the ϕ, ψ related dihedral parameters was performed to improve agreement with QM data for both the alanine dipeptide and tetrapeptide (Beachy et al., 1997), yielding PARM99. More recently, adjustments have been made to deal with the tendency of PARM99 to favor α-helical conformations. Modifications include alteration of the ϕ, ψ dihedral parameters and changes in the charge distribution for the entire protein force field. The dihedral parameters were changed in two studies to alter the conformational space being sampled in peptide simulations

(Garcia and Sanbonmatsu, 2002; Okur et al., 2003). In another study, the partial atomic charges were redetermined via RESP fits to B3LYP/cc-pVTZ//HF/6-31G* QM data (Duan et al., 2003). This was followed by readjustment of the ϕ, ψ-related dihedral parameters to reproduce QM maps of the alanine and glycine dipeptide maps obtained at the MP2/cc-pVTZ//HF/6-31G* level with a dielectric constant of 4, where the dielectric was selected to mimic the environment on the protein interior. While the efforts at additional optimization of the force fields are admirable, care must be taken that the changes are done in an orderly, well-defined manner. For example, as discussed above, alteration of the charges will impact the remaining parameters in the force field, possibly requiring their readjustment. In addition, the development of a collection of variants of a force field may become problematic in that difficulties in comparing results from different studies may arise.

2.6 Extended or United Atom Protein Force Fields

An alternative to the all-atom protein force fields is the extended or united atom force fields. Such force fields typically omit nonpolar hydrogen atoms to save computer time; these models dominated the early protein force fields. Examples of extended atom models include CHARMM PARAM19 (Neria et al., 1996), OPLS/UA (Jorgensen and Tirado-Rives, 1988), the early AMBER force fields (Weiner and Kollman, 1981; Weiner et al., 1984, 1986), and GROMOS87 and 96 (van Gunsteren et al., 1996). The GROMOS force field is still widely used in MD simulations that include explicit solvent representations. GROMOS96 has been subjected to tests in the condensed phase (Daura et al., 1996) and improved LJ parameters have recently been reported (Schuler et al., 2001). The other united atom force fields are primarily used for simulations on long time scales via the use of implicit solvent models, with the majority of these studies being based on PARAM19. PARAM19 can be used with several continuum solvation models including EEF1 (Lazaridis and Karplus, 1999), ACE (Schaefer et al., 2001), several GB models (Lee et al., 2002; Im et al., 2003; Lee et al., 2003), and a buried surface area model by Caflisch and co-workers (Ferrara et al., 2002). A summary of recent applications of both united and all-atom protein force fields combined with implicit solvent models has been presented (Feig and Brooks, 2004).

Several other force fields have been used for protein simulations, although they have not been used extensively. These include ENCAD (Levitt, 1990; Levitt et al., 1995), CEDAR (Ferro et al., 1980; Hermans et al., 1984), MMFF (Halgren, 1999) and CVFF (Ewig et al., 2001), among others. A more comprehensive list has been presented by Ponder and Case (Ponder and Case 2003). While these and other force fields may be used for protein simulations, it should be emphasized that they should be subjected to tests to ensure that they are appropriate for the problem under study.

Recently, the first reports of MD simulations of force fields that include explicit treatment of electronic polarizability (Halgren and Damm, 2001; Rick and Stuart, 2002) have appeared in the literature. These include studies using a fluctuating charge model (Patel and Brooks, 2004; Patel et al., 2004) and a point-dipole model (Kaminski et al., 2002; Harder et al., 2005). In addition, an MD simulation of DNA using a classical Drude oscillator to treat electronic polarizability has been published (Anisimov et al., 2005; Lamoureux et al., 2006) and this model is expected to be extended to proteins in the near future (A.D. MacKerell, Jr. and B. Roux, work in progress). The main advantage of polarizable force fields is the ability to more accurately model environments of different polarities, such as the polar environment on the surfaces of a protein versus the more hydrophobic interior of the protein. This improvement is at the cost of computer time, as treatment of the polarizability (Rick et al., 1995; Tuckerman and Martyna, 2000) can significantly increase the computational costs. In addition, the currently available first-generation polarizable force fields will probably require improvements as they are more rigorously tested on a wide variety of proteins.

2.7 Summary

A wide variety of empirical force fields for proteins are available and have been tested on a large number of peptides and proteins as well as on small molecule model systems. Currently, the all-atom force fields based on an additive model (e.g., no explicit treatment of electronic polarizability) are the most commonly used and have been shown to reproduce many types of experimental data in a wide variety of systems. While many protein simulations are performed in the presence of explicit solvent, improvements in implicit or continuum solvent models allow for protein simulations to be performed at a considerable savings of computational resources. While implicit models have been particularly useful in protein folding studies, it should be emphasized that in situations where individual water molecules play an essential role in protein structure and function, they will typically fail. With respect to the future, force fields that include explicit treatment of electronic polarization offer the potential of more accurately treating the wide range of environments in and around proteins. However, the current additive models will often be the method of choice when extended sampling of conformational space or time ranges is required.

Acknowledgments

Financial support from the NIH (GM51501 and GM 072558) and the University of Maryland, School of Pharmacy, Computer-Aided Drug Design Center is acknowledged.

Appendix

Web sites associated with commonly used protein force fields and simulation packages

CHARMM	*www.charmm.org* *www.pharmacy.umaryland.edu/faculty/amackere/* CHARMM also allows for simulations using AMBER, MMFF, and OPLS/AA force fields.
AMBER	*amber.scripps.edu/*
GROMOS	*www.igc.ethz.ch/gromos/*
OPLS	*zarbi.chem.yale.edu/software.html* *www.cs.sandia.gov/projects/towhee/forcefields/oplsaa.html*
Tinker	*dasher.wustl.edu/tinker/* Includes the CHARMM, AMBER, and OPLS force fields among others.
CVFF	*www.accelrys.com*
CEDAR	*femto.med.unc.edu/Hermans/jhermans.html*

References

Allen, F. H., S. Bellard, M. D. Brice, B. A. Cartwright, A. Doubleday, H. Higgs, T. Hummelink, B. G. Hummelink-Peters, O. Kennard, W. D. S. Motherwell, J. R. Rodgers, and D. G. Watson. 1979. The Cambridge Crystallographic Data Centre: Computer-based search, retrieval, analysis and display of information. *Acta Crystallogr. Sect. B* 35:2331–2339.

Allen, M. P., and D. J. Tildesley. 1989. *Computer Simulation of Liquids*. New York, Oxford University Press.

Anisimov, V. M., G. Lamoureux, I. V. Vorobyov, N. Huang, B. Roux, and A. D. MacKerell, Jr. 2005. Determination of electrostatic parameters for a polarizable force field based on the classical drude oscillator. *J. Chem. Theory Comput.* 1:153–168.

Banavali, N. K., and B. Roux. 2002. Atomic radii for continuum electrostatic calculations on nucleic acids. *J. Phys. Chem. B* 106:11026–11035.

Barth, E., and T. Schlick. 1998. Extrapolation versus impulse in multiple-timestepping schemes. II. Linear analysis and applications to Newtonian and Langevin dynamics. *J. Chem. Phys.* 109:1633–1642.

Bayly, C. I., P. Cieplak, W. D. Cornell, and P. A. Kollman. 1993. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: The RESP model. *J. Phys. Chem.* 97:10269–10280.

Beachy, M. D., D. Chasman, R. B. Murphy, T. A. Halgren, and R. A. Friesner. 1997. Accurate ab initio quantum chemical determination of the relative energetics of peptide conformations and assessment of empirical force fields. *J. Am. Chem. Soc.* 119:5908–5920.

Beglov, D., and B. Roux. 1994. Finite representation of an infinite bulk system: Solvent boundary potential for computer simulations. *J. Chem. Phys.* 100:9050–9063.

Berendsen, H. J. C., J. R. Grigera, and T. P. Straatsma. 1987. The missing term in effective pair potentials. *J. Phys. Chem.* 91:6269–6271.

Berman, H. M., T. Battistuz, T. N. Bhat, W. F. Bluhm, P. E. Bourne, K. Burkhardt, Z. Feng, G. L. Gilliland, L. Iype, S. Jain, P. Fagan, J. Marvin, D. Padilla, V. Ravichandran, B. Schneider, N. Thanki, H. Weissig, J. D. Westbrook, and C. Zardecki. 2002. The protein data bank. *Acta. Crystallogr. D. Biol. Crystallogr.* 58:899–907.

Bishop, T. C., R. D. Skeel, and K. Schulten. 1997. Difficulties with multiple time stepping and fast multipole algorithm in molecular dynamics. *J. Comput. Chem.* 18:1785–1791.

Blondel, A., and M. Karplus. 1996. New formulation of derivatives of torsion angles and improper torsion angles in molecular mechanics: Elimination of singularities. *J. Comput. Chem.* 17:1132–1141.

Brooks, B. R., R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. 1983. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* 4:187–217.

Buck, M., S. Bonnet, R. W. Pastor, and A. D. MacKerell, Jr. 2006. Importance of the CMAP correction to the CHARMM22 Protein Force Field: Dynamics of hen lysozyme. *Biophys. J.* 90:L36–L38.

Buckingham, A. D., and P. W. Fowler. 1985. A model for the geometries of van der Waals complexes. *Can. J. Chem.* 63:2018.

Burkert, U., and N. L. Allinger. 1982. *Molecular Mechanics*. Washington, DC, American Chemical Society.

Bush, B. L., C. I. Bayly, and T. A. Halgren. 1999. Consensus bond-charge increments fitted to electrostatic potential or field of many compounds: Application of MMFF94 training set. *J. Comput. Chem.* 20:1495–1516.

Chalasinski, G., and M. M. Szczesniak. 1994. Origins of structure and energetics of van der Waals clusters from ab initio calculations. *Chem. Rev.* 94:1723–1765.

Chen, I.-J., D. Yin, and A. D. MacKerell, Jr. 2002. Combined ab initio/empirical optimization of Lennard-Jones parameters for polar neutral compounds. *J. Comput. Chem.* 23:199–213.

Chirlian, L. E., and M. M. Francl. 1987. Atomic charges derived from electrostatic potentials: A detailed study. *J. Comput. Chem.* 8:894–905.

Cieplak, P., W. D. Cornell, C. I. Bayly, and P. K. Kollman. 1995. Application of the multimolecule and multiconformational RESP methodlogy to biopolymers: Charge derivation for DNA, RNA, and proteins. *J. Comput. Chem.* 16:1357–1377.

Cornell, W. D., P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. 1995. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* 117:5179–5197.

Darden, T. 2001. Treatment of long-range forces and potentials. In *Computational Biochemistry and Biophysics* (O. M. Becker, A. D. MacKerell, Jr., B. Roux, and M. Watanabe, Eds.). New York, Dekker, pp. 91–114.

Daura, X., P. H. Hünenberger, A. E. Mark, E. Querol, F. X. Avilés, and W. F. van Gunsteren. 1996. Free energies of transfer of Trp analogs from chloroform to water: Comparison of theory and experiment and the importance of adequate treatment of electrostatics and internal interactions. *J. Am. Chem. Soc.* 118:6285– 6294.

Davis, M. E., and J. A. McCammon. 1990. Electrostatics in biomolecular structure and dynamics. *Chem. Rev.* 90:509–521.

Deng, Y., and B. Roux. 2004. Hydration of amino acid side chains: Non-polar and electrostatic contributions calculated from staged molecular dynamics free energy simulations with explicit water molecules. *J. Phys. Chem. B* 108:16567–16576.

Derreumaux, P., and G. Vergoten. 1995. A new spectroscopic molecular mechanics force field. Parameters for proteins. *J. Chem. Phys.* 102:8586–8605.

Duan, Y., C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Ceiplak, R. Luo, T. Lee, J. Caldwell, J. Wang, and P. Kollman. 2003. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* 24:1999–2012.

Dunbrack, R. L., Jr. 2002. Culledpdb: Non-redundant set of protein sidechains from the PDB. Philadelphia.

Dunbrack, R. L., Jr., and F. E. Cohen. 1997. Bayesian statistical analysis of protein sidechain rotamer preferences. *Protein Sci.* 6:1661–1681.

Elber, R., and M. Karplus. 1990. Enhanced sampling in molecular dynamics: Use of the time-dependent hartree approximation for a simulation of carbon monoxide diffusion through myoglobin. *J. Am. Chem. Soc.* 112:9161–9175.

Ewald, P. P. 1921. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Ann. Phys.* 64:253–287.

Ewig, C. S., R. Berry, U. Dinur, J.-R. Hill, M.-J. Hwang, H. Li, C. Liang, J. Maple, Z. Peng, T. P. Stockfisch, T. S. Thacher, L. Yan, X. Ni, and A. T. Hagler. 2001. Derivation of class II force fields. VIII. Derivation of a general quantum mechanical force field for organic compounds. *J. Comput. Chem.* 22:1782–1800.

Feig, M., and C. L. Brooks III. 2004. Recent advances in the development and application of implicit solvent models in biomolecular simulations. *Curr. Opin. Struct. Biol.* 14:217–224.

Feller, S. E., K. Gawrisch, and A. D. MacKerell, Jr. 2002. Polyunsaturated fatty acids in lipid bilayers: Intrinsic and environmental contributions to their unique physical properties. *J. Am. Chem. Soc.* 124:318–326.

Feller, S. E., R. W. Pastor, A. Rojnuckarin, S. Bogusz, and B. R. Brooks. 1996. Effect of electrostatic force truncation on interfacial and transport properties of water. *J. Phys. Chem.* 100:17011–17020.

Feller, S. E., Y. Zhang, R. W. Pastor, and R. W. Brooks. 1995. Constant pressure molecular dynamics simulation: The Langevin piston method. *J. Chem. Phys.* 103:4613–4621.

Ferrara, P., J. Apostolakis, and A. Caflisch. 2002. Evaluation of a fast implicit solvent model for molecular dynamics simulations. *Proteins* 46:24–33.

Ferrara, P., H. Gohlke, D. J. Price, G. Klebe, and C. L. I. Brooks. 2004. Assessing scoring functions for protein–ligand interactions. *J. Med. Chem.* 47:3032–3047.

Ferro, D. R., J. E. McQueen, J. T. McCown, and J. Hermans. 1980. Energy minimization of rubredoxin. *J. Mol. Biol.* 136:1–18.

Florián, J., and A. Warshel. 1997. Langevin dipoles model for ab initio calculations of chemical processes in solution: Parameterization and application to hydration free energies of neutral and ionic solutes and conformational analysis in aqueous sdolution. *J. Phys. Chem. B* 101:5583–5595.

Foloppe, N., and A. D. MacKerell, Jr. 2000. All-atom empirical force field for nucleic acids: 1. Parameter optimization based on small molecule and condensed phase macromolecular target data. *J. Comput. Chem.* 21:86–104.

Francl, M. M., C. Carey, L. E. Chirlian, and D. M. Gange. 1996. Charge fit to electrostatic potentials. II. Can atomic charges be unambiguously fit to electrostatic potentials? *J. Comput. Chem.* 17:367–383.

Freedberg, D. I., R. M. Venable, A. Rossi, T. E. Bull, and R. W. Pastor. 2004. Discriminating the helical forms of peptides by NMR and molecular dynamics simulations. *J. Am. Chem. Soc.* 126:10478–10484.

Gallicchio, E., and R. M. Levy. 2004. AGBNP: An analytic implicit solvent model suitable for molecular dynamics simulations and high-resolution modeling. *J. Comput. Chem.* 25:479–499.

Gallicchio, E., L. Y. Zhang, and R. M. Levy. 2003. The SGB/NP hydration free energy model based on the surface generalized Born solvent reaction field and novel nonpolar hydration free energy estimators. *J. Comput. Chem.* 23:517–529.

Garcia, A. E., and K. Y. Sanbonmatsu. 2002. α-Helical stabilization by side chain sheilding of backbone hydrogen bonds. *Proc. Natl. Acad. Sci. USA* 99:2782–2787.

Gilson, M. K., H. S. Gilson, and M. J. Potter. 2003. Fast assignment of accurate partial atomic charges: An electronegativity equilization method that accounts for alternate resonance forms. *J. Chem. Inf. Comput. Sci.* 43:1982–1997.

Gilson, M. K., and B. Honig. 1988. Calculation of the total electrostatic energy of a macromolecular system: Solvation energies, binding energies, and conformational analysis. *Proteins* 4:7–18.

Glättli, A., X. Daura, and W. F. van Gunsteren. 2003. A novel approach for designing simple point charge models for liquid water with three interaction sites. *J. Comput. Chem.* 24:1087–1096.

Gohlke, H., C. Kiel, and D. Case. 2003. Protein–protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *J. Mol. Biol* 330:891–913.

Habtemariam, B., V. M. Anisimov, and A. D. MacKerell, Jr. 2005. Cooperative binding of DNA and CBFβ to the runt domain of the CBFα studied via MD simulations. *Nucleic Acids Res.* 33:4212–4222.

Halgren, T. A. 1992. Representation of van der Waals (vdW) interactions in molecular mechanics force fields: Potential form, combination rules, and vdW parameters. *J. Am. Chem. Soc.* 114:7827–7843.

Halgren, T. A. 1996a. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* 17: 490–519.

Halgren, T. A. 1996b. Merck molecular force field. II. MMFF94 van der Waals and electrostatic parameters for intermolecular interactions. *J. Comput. Chem.* 17:520–552.

Halgren, T. A. 1996c. Merck molecular force field. III. Molecular geometries and vibrational frequencies for MMFF94. *J. Comput. Chem.* 17:553–586.

Halgren, T. A. 1999. MMFF VII. Characterization of MMFF94, MMFF94s, and other widely available force fields for conformational energies and for intermolecular-interaction energies and geometries. *J. Comput. Chem.* 20:730–748.

Halgren, T. A., and W. Damm. 2001. Polarizable force fields. *Curr. Opin. Struct. Biol.* 11:236–242.

Hansmann, U. H. E. 1997. Parallel tempering algorithm for conformational studies of biological molecules. *Chem. Phys. Lett.* 281:140–150.

Harder, E., B. Kim, R. A. Friesner, and B. J. Berne. 2005. Efficient simulation method for polarizable protein force fields: Application to the simulation of BPTI in liquid water. *J. Chem. Theory Comput.* 1:169–180.

Head-Gordon, T., M. Head-Gordon, M. J. Frisch, C. L. Brooks, and J. A. Pople. 1991. Theoretical study of blocked glycine and alanine peptide analogues. *J. Am. Chem. Soc.* 113:5989–5997.

Henchman, R. H., and J. W. Essex. 1999. Generation of OPLS-like charges from molecular electrostatic potential using restraints. *J. Comput. Chem.* 20:483–498.

Hermans, J., H. J. C. Berendsen, W. F. van Gunsteren, and J. P. M. Postma. 1984. A consistent empirical potential for water–protein interactions. *Biopolymers* 23:1513–1518.

Honig, B. 1993. Macroscopic models of aqueous solutions: Biological and chemical applications. *J. Phys. Chem.* 97:1101.

Hu, H., M. Elstner, and J. Hermans. 2003. Comparison of a QM/MM force field and molecular mechanics force fields in simulations of alanine and glycine "dipeptides" (Ace-Ala-Nme and Ace-Gly-Nme) in water in relation to the problem of modeling the unfolded peptide backbone in solution. *Proteins Struct Funct. Genet.* 50:451–463.

Huang, N., and A. D. MacKerell, Jr. 2002. An *ab initio* quantum mechanical study of hydrogen-bonded complexes of biological interest. *J. Phys. Chem. B* 106:7820–7827.

Im, W., S. Bernéche, and B. Roux. 2001. Generalized solvent boundary potential for computer simulations. *J. Chem. Phys.* 114:2924–2937.

Im, W., M. S. Lee, and C. L. Brooks III. 2003. Generalized Born model with a simple smoothing function. *J. Comput. Chem.* 24:1691–702.

Jakalian, A., B. L. Bush, D. B. Jack, and C. I. Bayly. 2000. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method. *J. Comput. Chem.* 21:132–146.

Jayaram, B., K. J. McConnell, S. B. Dixit, and D. L. Beveridge. 2002. Free-energy component analysis of 40 protein–DNA complexes: A consensus view on the thermodynamics of binding at the macromolecular level. *J. Comput. Chem.* 23:1–14.

Jayaram, B., D. Sprous, and D. L. Beveridge. 1998. Solvation free energy of biomacromolecules: Parameters for a modified generalized Born model consistent with the AMBER force field. *J. Phys. Chem. B* 102: 9571–9576.

Jorgensen, W. L. 1984. Optimized intermolecular potential functions for liquid hydrocarbons. *J. Am. Chem. Soc.* 106:6638–6646.

Jorgensen, W. L. 1986. Optimized intermolecular potential functions for liquid alcohols. *J. Phys. Chem.* 90:1276–1284.

Jorgensen, W. L., J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935.

Jorgensen, W. L., D. S. Maxwell, and J. Tirado-Rives. 1996. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* 118:11225–11236.

Jorgensen, W. L., and J. Tirado-Rives. 1988. The OPLS potential function for proteins. Energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* 110:1657–1666.

Kaminski, G., R. A. Friesner, J. Tirado-Rives, and W. L. Jorgensen. 2001. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B* 105:6474–6487.

Kaminski, G. A., H. A. Stern, B. J. Berne, R. A. Friesner, Y. X. Cao, R. B. Murphy, R. Zhou, and T. A. Halgren. 2002. Development of a polarizable force field for proteins via ab initio quantum chemistry: First generation model and gas phase tests. *J. Comput. Chem.* 23: 1515–1531.

Kim, K., and R. A. Friesner. 1997. Hydrogen bonding between amino acid backbone and side chain analogues: A high-level ab initio study. *J. Am. Chem. Soc.* 119:12952–12961.

Klauda, J. B., B. R. Brooks, A. D. MacKerell, Jr., R. M. Venable, and R. W. Pastor. 2005. 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* 109:5300–5311.

Kollman, P. A., I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. A. Case, and T. E. Cheatham III. 2000. Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Acc. Chem. Res.* 33:889–897.

Lague, P., R. W. Pastor, and B. R. Brooks. 2004. A pressure-based long-range correction for Lennard Jones interactions in molecular dynamics simulations: Application to alkanes and interfaces. *J. Phys. Chem. B* 108:363–368.

Laio, A., J. VandeVondele, and U. Rothlisberger. 2002. D-RESP: Dynamically generated electrostatic potential derived charges from quantum mechanics/molecular mechanics simulations. *J. Phys. Chem. B* 106:7300–7307.

Lamoureux, G., E. Harder, I. V. Vorobyov, B. Roux, and A. D. MacKerell, Jr. 2005. A polarizable model of water for molecular dynamics simulations of biomolecules. *Chem. Phys. Lett.* 418:241–245.

Lazaridis, T., and M. Karplus. 1999. Effective energy function for proteins in solution. *Proteins* 35:133–152.

Lee, M. S., M. Feig, F. R. Salsbury, Jr., and C. L. Brooks III. 2003. New analytical approximation of the standard molecular volume definition and its application to generalized Born calculations. *J. Comput. Chem.* 24:1348–1356.

Lee, M. S., F. R. Salsbury, Jr., and C. L. Brooks III. 2002. Novel generalized Born methods. *J. Chem. Phys.* 116:10606–10614.

Levitt, M. 1990. ENCAD—Energy Calculations and Dynamics. Stanford, CA and Rehovot, Israel, Molecular Applications Group.

Levitt, M., M. Hirshberg, R. Sharon, and V. Daggett. 1995. Potential energy function and parameters for simulations of the molecular dynamics of proteins and nucleic acids in solution. *Comput. Phys. Commun.* 91:215–231.

Levitt, M., M. Hirshberg, R. Sharon, K. E. Laidig, and V. Daggett. 1997. Calibration and testing of a water model for simulation of the molecular dynamics of proteins and nucleic acids in solution. *J. Phys. Chem. B* 101:5051–5061.

Lii, J.-L., and N. L. Allinger. 1991. The MM3 force field for amides, polypeptides and proteins. *J. Comput. Chem.* 12:186–199.

MacCallum, J. L., and P. Tieleman. 2003. Calculation of the water–cyclohexane transfer free energies of amino acid side-chain analogs using the OPLS all-atom force field. *J. Comput. Chem.* 24:1930–1935.

Macias, A. T., and A. D. MacKerell, Jr. 2005. CH/pi interactions involving aromatic amino acids: Refinement of the CHARMM tryptophan force field. *J. Comput. Chem.* 26:1452–1463.

MacKerell, A. D., Jr. 2001. Atomistic models and force fields. In *Computational Biochemistry and Biophysics* (O. M. Becker, A. D. MacKerell, Jr., B. Roux, and M. Watanabe, Eds.). New York, Dekker, pp. 7–38.

MacKerell, A. D., Jr. 2004. Empirical force fields for biological macromolecules: Overview and issues. *J. Comput. Chem.* 25:1584–1604.

MacKerell, A. D., Jr., D. Bashford, M. Bellott, R. L. Dunbrack, Jr., J. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, I. Reiher, W. E., B. Roux, M. Schlenkrich, J. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus. 1998a. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* 102:3586–3616.

MacKerell, A. D., Jr., B. Brooks, C. L. Brooks III, L. Nilsson, B. Roux, Y. Won, and M. Karplus. 1998b. CHARMM: The energy function and its paramerization with an overview of the program. In *Encyclopedia of Computational Chemistry* (P. v. R. Schleyer et al., Eds.) Chichester, John Wiley & Sons. pp. 271–277.

MacKerell, A. D., Jr., M. Feig, and C. L. Brooks III. 2004a. Accurate treatment of protein backbone conformational energetics in empirical force fields. *J. Am. Chem. Soc.* 126:698–699.

MacKerell, A. D., Jr., M. Feig, and C. L. Brooks III. 2004b. 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.* 25:1400–1415.

MacKerell, A. D., Jr., J. Wiórkiewicz-Kuczera, and M. Karplus. 1995. An all-atom empirical energy function for the simulation of nucleic acids. *J. Am. Chem. Soc.* 117:11946–11975.

Mahoney, M. W., and W. L. Jorgensen. 2000. A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *J. Chem. Phys.* 112:8910–8922.

Martyna, G. J., D. J. Tobias, and M. L. Klein. 1994. Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* 101:4177–4189.

Mayo, S. L., B. D. Olafson, and W. A Goddard III. 1990. 'DREIDING: A generic force field for molecular simulations.' *J. Phys. Chem.* 94:8897–8909.

McQuarrie, D. A. 1976. *Statistical Mechanics*. New York, Harper & Row.

Merz, K. M. 1992. Analysis of a large data base of electrostatic potential derived atomic charges. *J. Comput. Chem.* 13:749–767.

Neria, E., S. Fischer, and M. Karplus. 1996. Simulation of activation free energies in molecular systems. *J. Chem. Phys.* 105:1902–1919.

Nina, M., D. Beglov, and B. Roux. 1997. Atomic radii for continuum electrostatics calculation based on molecular dynamics free energy simulations. *J. Phys. Chem. B* 101:5239–5248.

Nymeyer, H., S. Gnanakaran, and A. E. Garcia. 2004. Atomic simulations of protein folding, using the replica exchange algorithm. *Methods Enzymol.* 383:119–149.

Okur, A., B. Strockbine, V. Hornak, and C. Simmerling. 2003. Using PC clusters to evaluate the transferability of molecular mechanics force fields for proteins. *J. Comput. Chem.* 24:21–31.

Ono, S., M. Kuroda, J. Higo, N. Nakajima, and H. Nakamura. 2002. Calibration of force-field dependency in free energy landscapes of peptide conformations by quantum chemical calculations. *J. Comput. Chem.* 23:470–476.

Onufriev, A., D. Bashford, and D. A. Case. 2000. Modification of the generalized born model suitable for macromolecules. *J. Phys. Chem. B* 104:3712–3720.

Palmo, K., B. Mannfors, N. G. Mirkin, and S. Krimm. 2003. Potential energy functions: From consistent force fields to spectroscopically determined polarizable force fields. *Biopolymers* 68:383–394.

Patel, S., and C. L. Brooks III. 2004. CHARMM fluctuating charge force field for proteins: I Parameterization and application to bulk organic liquid simulations. *J. Comput. Chem.* 25:1–15.

Patel, S., A. D. MacKerell, Jr., and C. L. Brooks III. 2004. CHARMM fluctuating charge force field for proteins: II Protein/solvent properties from molecular dynamics simulations using a nonadditive electrostatic model. *J. Comput. Chem.* 25:1504–1514.

Ponder, J. W., and D. A. Case. 2003. Force fields for protein simulations. *Adv. Protein Chem.* 66:27–85.

Price, D. J., and C. L. Brooks III. 2002. Modern protein force fields behave comparably in molecular dynamics simulations. *J. Comput. Chem.* 23:1045–1057.

Qui, D., P. S. Shenkin, F. P. Hollinger, and W. C. Still. 1997. The GB/SA continuum model for solvation. A fast analytical method for the calculation of approximate Born radii. *J. Phys. Chem. A* 101:3005–3014.

Ramachandran, G. N., C. Ramakrishnan, and V. Sasisekharan. 1963. Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* 7:95–99.

Rappé, A. K., C. J. Colwell, W. A. Goddard III, and W. M. Skiff. 1992. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J. Am. Chem. Soc.* 114:10024–10035.

Reiher, W. E. 1985. Theoretical Studies of Hydrogen Bonding Ph.D. thesis, Harvard University.

Rick, S. W., and S. J. Stuart. 2002. Potentials and algorithms for incorporating polarizability in computer simulations. *Rev. Comput. Chem.* 18:89–146.

Rick, S. W., S. J. Stuart, J. S. Bader, and B. J. Berne. 1995. Fluctuating charge force fields for aqueous solutions. *J. Mol. Liq.* 65/66:31–40.

Rizzo, R. C., and W. L. Jorgensen. 1999. OPLS all-atom model for amines: Resolution of the amine hydration problem. *J. Am. Chem. Soc.* 121:4827–4836.

Schaefer, M., C. Bartels, F. LeClerc, and M. Karplus. 2001. Effective atom volumes for implicit solvent models: Comparison between Voronoi volumes and minimum fluctuation volumes. *J. Comput. Chem.* 22:1857–1879.

Schaefer, M., and M. Karplus. 1996. A comprehensive analytical treatment of continuum electrostatics. *J. Phys. Chem.* 100:1578–1599.

Schuler, L. D., X. Daura, and W. F. van Gunsteren. 2001. An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase. *J. Comput. Chem.* 22:1205–1218.

Shirts, M. R., J. W. Pitera, W. C. Swope, and V. S. Pande. 2003. Extremely precise free energy calculations of amino acid side chain analogs: Comparison of common molecular mechanics force fields for proteins. *J. Chem. Phys.* 119:5740–5761.

Simmerling, C., T. Fox, and P. A. Kollman. 1998. Use of locally enhanced sampling in free energy calculations: Testing and application to the $\alpha \rightarrow \beta$ anomerization of glucose. *J. Am. Chem. Soc.* 120:5771–5782.

Singh, U. C., and P. A. Kollman. 1984. An approach to computing electrostatic charges for molecules. *J. Comput. Chem.* 5:129–145.

Steinbach, P. J. 2004. Exploring peptide energy landscapes: A test of force fields and implicit solvent models. *Proteins* 57:665–677.

Still, W. C., A. Tempczyk, R. C. Hawley, and T. Hendrickson. 1990. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* 112:6127–6129.

Sugita, Y., and Y. Okamoto. 1999. Replica-exchange molecular dynamics methods for protein folding. *Chem. Phys. Lett.* 314:141–151.

Sun, H. 1998. COMPASS: An ab inito force-field optimized for condensed-phase applications-overview with details on alkane and benzene compounds. *J. Phys. Chem. B* 102:7338–7364.

Tuckerman, M., B. J. Berne, and G. J. Martyna. 1992. Reversible multiple time scale molecular dynamics. *J. Chem. Phys.* 97:1990–2001.

Tuckerman, M. E., and G. J. Martyna. 2000. Understanding modern molecular dynamics: Techniques and applications. *J. Phys. Chem. B* 104:159–178.

van Gunsteren, W. F. 1987. GROMOS. Groningen Molecular Simulation Program Package. Groningen, University of Groningen.

van Gunsteren, W. F., S. R. Billeter, A. A. Eising, P. H. Hünenberger, P. Krüger, A. E. Mark, W. R. P. Scott, and I. G. Tironi. 1996. *Biomolecular Simulation: The GROMOS96 Manual and User Guide*. Zürich, BIOMOS b.v.

Vargas, R., J. Garza, B. P. Hay, and D. A. Dixon. 2002. Conformational study of the alanine dipeptide at the MP2 and DFT levels. *J. Phys. Chem. A* 106:3213–3218.

Villa, A., and A. E. Mark. 2002. Calculation of the free energy of solvation for neutral analogs of amino acid side chains. *J. Comput. Chem.* 23:548–553.

Wang, J., and P. A. Kollman. 2001. Automatic parameterization of force field by systematic search and genetic algorithms. *J. Comput. Chem.* 22:1219–1228.

Warshel, A., and S. Lifson. 1970. Consistent force field calculations. II. Crystal structures, sublimation energies, molecular and lattice vibrations, molecular conformations, and enthalpy of alkanes. *J. Chem. Phys.* 53:582–594.

Weiner, P. K., and P. A. Kollman. 1981. AMBER: Assisted Model Building with Energy Refinement. A general program for modeling molecules and their interactions. *J. Comput. Chem.* 2:287–303.

Weiner, S. J., P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta, and P. Weiner. 1984. A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* 106:765–784.

Weiner, S. J., P. A. Kollman, D. T. Nguyen, and D. A. Case. 1986. An all atom force field for simulations of proteins and nucleic acids. *J. Comput. Chem.* 7:230–252.

Yin, D., and A. D. MacKerell, Jr. 1996. Ab initio calculations on the use of helium and neon as probes of the van der Waals surfaces of molecules. *J. Phys. Chem.* 100:2588–2596.

Yin, D., and A. D. MacKerell, Jr. 1998. Combined ab initio/empirical approach for the optimization of Lennard-Jones parameters. *J. Comput. Chem.* 19:334–348.

Zhang, L. Y., E. Gallicchio, R. A. Friesner, and R. M. Levy. 2001. Solvent models for protein–ligand binding: Comparison of implicit solvent Poisson and surface generalized Born models with explicit solvent simulations. *J. Comput. Chem.* 22:591–607.