Empirical Force Fields for Biological Macromolecules: Overview and Issues

ALEXANDER D. MACKERELL, JR.
Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, 20 Penn Street, Baltimore, Maryland 21201

Received 12 April 2004; Accepted 2 May 2004
DOI 10.1002/jcc.20082
Published online in Wiley InterScience (www.interscience.wiley.com).

Abstract: Empirical force field-based studies of biological macromolecules are becoming a common tool for investigating their structure–activity relationships at an atomic level of detail. Such studies facilitate interpretation of experimental data and allow for information not readily accessible to experimental methods to be obtained. A large part of the success of empirical force field-based methods is the quality of the force fields combined with the algorithmic advances that allow for more accurate reproduction of experimental observables. Presented is an overview of the issues associated with the development and application of empirical force fields to biomolecular systems. This is followed by a summary of the force fields commonly applied to the different classes of biomolecules; proteins, nucleic acids, lipids, and carbohydrates. In addition, issues associated with computational studies on "heterogeneous" biomolecular systems and the transferability of force fields to a wide range of organic molecules of pharmacological interest are discussed.

 J Comput Chem 25: 1584–1604, 2004

Key words: molecular dynamics; molecular mechanics; CHARMM; AMBER; OPLS; GROMOS

Introduction

Empirical force field-based methods represent a major tool for the application of theoretical approaches to investigate structure–activity relationships in biological systems.[1,2] Currently, MD simulations of systems of over 100,000 atoms or more can be performed for time periods in the nanosecond regimen and beyond. Accessing systems of such sizes and time scales is based on ever increasing computer power, as exemplified in Moore's law,[3] combined with algorithmic developments allowing for efficient use of highly parallel computers. Central to the continued and future success of these methods is their accuracy, as judged by their ability to reproduce experimentally accessible properties. Algorithmic improvements, including more rigorous approaches to treat long range nonbond interactions, including Ewald[4,5] and other boundary methods,[6–8] and improved integrators,[9–13] that allow for proper thermodynamic ensembles to be simulated, have made a major contribution towards improved accuracy. However, the quality of the force fields themselves, combined with their proper implementation, may be considered the most important determinant of the accuracy of these empirical methods.

In the present review we will give a summary of empirical force fields, with emphasis on those commonly used for empirical studies of biomolecules. This will include presentation of the form of the potential energy function commonly used in biomolecular force fields, considerations as to the proper implementation of biomolecular force fields and the different types of solvation models in use, including both explicit and implicit models. Following this will be overviews of the force fields in use for the different classes of biomolecules; proteins, nucleic acids, lipids, and carbohydrates, as well as considerations for simulations of heterogeneous biomolecular systems that include more than one class of biomolecule and to the applicability of force fields to the wide-range of organic molecules of pharmaceutical interest.

The ultimate goal of the present overview is facilitation of the application of empirical force fields to biological systems, supplying users with the information required to select the most appropriate force field for the system under study. However, it should be stated that the author's intimate relationship with the CHARMM force fields[14,15] may lead to some bias in the present article. Thus, when reading this review, as when using a force field, the source should be considered when analyzing the outcome.

Correspondence to: A.D. MacKerell, Jr.; e-mail: alex@outerbanks.umaryland.edu

Contract/grant sponsor: the NIH; contract/grant number: GM51501

Conttact/grant sponsor: the University of Maryland, School of Pharmacy Computer-Aided Drug Design Center

Potential Energy Functions

The core of any force field is the potential energy function used to describe the relationship of the structure, $\vec{R}$, to the energy, U, of the system of interest. However, a potential energy function alone does not make a force field. It is the combination of the potential energy function with the parameters used in that function, as described below, that yield a force field. In the remainder of this section details of the potential energy functions used in biomolecular force fields will be presented followed by issues associated with the parameters themselves.

Equation (1) shows an example of what is referred to as a Class I additive potential energy function. The form of this function is similar to that applied in early force fields[16] and is still the form used in biomolecular force fields commonly in use as of the writing of this article. As is evident, eq. (1) is comprised of a collection of simple functions

$$U(\vec{R}) = \sum_{bonds} K_b(b - b_0)^2 + \sum_{angles} K_\theta(\theta - \theta_0)^2 + \sum_{dihedral} \times K_\chi\Big(1 + \cos(n\chi - \delta)\Big) + \sum_{impropers} K_{imp}(\varphi - \varphi_0)^2 + \sum_{nonbond} \left(\varepsilon_{ij}\left[\left(\frac{R\min_{ij}}{r_{ij}}\right)^{12} - \left(\frac{R\min_{ij}}{r_{ij}}\right)^{6}\right]\right) + \frac{q_i q_j}{\varepsilon r_{ij}} \quad (1)$$

to represent a minimal set of forces that can describe molecular structures. Bonds, angles, and out-of-plane distortions (improper dihedral angles) are treated harmonically and dihedral or torsional rotations are described by a sinusoidal term. Interactions between atoms use a Lennard–Jones (LJ) 6–12 term to describe the atom–atom repulsion and dispersion interactions combined with electrostatics treated via a Coulombic term. This simple form for the potential energy function is necessitated by computational considerations, allowing studies on systems of 100,000 or more atoms. However, via proper adjustment of the parameters in eq. (1), this or similar functional forms can describe biological molecules with high accuracy.

In eq. (1), b is the bond length; θ, is the valence angle; χ, is the dihedral or torsion angle; φ, is the improper angle; and r_{ij} is the distance between atoms i and j. Parameters, the terms that represent the actual force field, 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, respectively; the dihedral force constant, multiplicity and phase angle, K_χ, n, and δ, respectively; and the improper force constant and equilibrium improper angle, K_φ and φ_0, respectively. Collectively, these represent the internal or intramolecular parameters. Nonbonded parameters 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 van der Waals (vdW) interactions. These terms are also referred to as the interaction or external parameters. 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 (see below). The dielectric constant, ε, is typically set to 1, corresponding to the permittivity of vacuum, in calculations that incorporate explicit solvent representations; alternative methods to treat the solvent environment are discussed briefly below. The terms contributing to the energy in eq. (1) are common to the majority of currently used biomolecular force fields, including CHARMM,[14,15] AMBER,[17] GROMOS,[18] OPLS,[19] among others.

Beyond eq. (1) are additional or alternate terms for both the internal and external aspects of the force field. A number of force fields, often referred to as Class II force fields, include higher order terms to treat the bond and valence angle terms and/or cross terms between, for example, bonds and valence angles or valence angles and dihedrals.[20–25] These terms increase the accuracy of force fields to treat conformational energies especially at geometries significantly far from the minimum-energy or equilibrium values. They also facilitate the accurate treatment of vibrational spectra, although Class I force fields yield good quality spectra when proper optimization of the parameters is undertaken. Other alternatives include the use of a Morse function for bonds; such a function allows for bond breaking in an empirical force field, and a cosine-based angle term that is well behaved for near-linear valence angles.[26,27] With respect to the dihedral term a recent improvement avoids singularities associated with derivatives of torsion angle cosines and allows for application of any value of the phase[28] and, more recently, was the introduction of a two-dimensional (2D) grid-based dihedral energy correction map[29,30] that allows for any 2D dihedral surface [e.g., a quantum mechanical (QM) ϕ,ψ surface of the alanine dipeptide] to be reproduced nearly exactly by the force field (see below). In general, for biomolecular simulations performed in the vicinity of room temperature, the Class I force fields adequately treat both the intramolecular distortions, including relative conformational energies associated with large structural changes, which may occur in biomolecules.

Alternate forms of the nonbonded portion of potential energy functions involve both alternate treatment of vdW interactions and the electrostatics. The three primary alternatives to the LJ 6–12 term included in eq. (1) are designed to "soften" the repulsive wall associated with Pauli exclusion. For example, the Buckingham potential[31] uses an exponential term to treat repulsion while a buffered 14–7 term is used in the MMFF force field.[32] 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.[33] 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.

Currently, the majority of biomolecular force fields of general applicability treat the electrostatic interactions using the Coulombic term included in eq. (1). This model uses static, partial atomic charges [i.e., q_i in eq. (1)] and is often referred to as an additive force field. Additive force fields, via the combination of the Coulombic and LJ terms, have been shown to treat hydrogen bonding with reasonable accuracy, including the angular dependencies.[34] Accordingly, the majority of force fields no longer include explicit terms for hydrogen bonding as was done in earlier force fields,[14,35] although exceptions exist.[20] Notable is that the current electrostatic models do not explicitly treat electronic polarizability. Instead, polarizability is included implicitly by choosing partial atomic charges that overestimate molecular dipoles.[36] This overestimation is designed to approximate electrostatic interactions that occur in

the aqueous, condensed phase environment common to biomolecules. Using this approximation for the polarizability, additive force fields have been shown to do quite well in treating both structure and energetics in the condensed phase. For example, accurate heats of vaporization and molecular volumes,[37–41] free energies of solvation,[42–45] and crystal heats of sublimation and lattice parameters[46–48] have been reported. Although use of the additive models has been shown to reproduce various experimental properties for a number of systems, including biomolecular systems, recent free energy of solvation calculations[49] indicate that improvements in the models can be made; such improvements may be performed in the context of the current form of the potential energy function or in the context of an extended function that contains explicit treatment of electronic polarizability (see below). An example of an inherent limitation in current additive models is their inability to simultaneously treat molecules in environments of significantly differing polar character with high accuracy. A simple case is the need to significantly overestimate the interaction energy of the gas phase water dimer in order to accurately treat the pure solvent (see below).[42,50]

Electronic Polarizability

The explicit inclusion of electronic polarization will represent the next significant development in the treatment of nonbonded interactions in biomolecular force fields,[51,52] which are referred to as polarizable or nonadditive force fields. Methods to treat polarizability have recently been reviewed.[53] Briefly, electronic polarizability may be included in a potential energy function as an extension of eq. (1) with a term that describes the energy associated with polarization of the charge distribution, U_{pol}, which is typically determined via

$$U_{\text{pol}} = \frac{1}{2} \sum_i \mu_i E_i \tag{2}$$

where μ_i is the dipole moment of atom i and E_i is the electrostatic field at atom i. Solving eq. (2) is generally performed iteratively, allowing for the change in E_i as a function of μ_i to converge. The dipole moment may be expressed in terms of the polarizability of an atom, α_i, where $\mu_i = \alpha_i E_i$, with the polarizabilities being parameters that are assigned based on atom type. There are multiple methods to explicitly treat polarizability, the most common methods being induced dipole models,[54–58] fluctuating charge models,[59–65] or a combination of those methods.[66,67] Induced dipole models generally apply an isotropic dipole moment to each atom, i, using a dipole field tensor, T_{ij}, to treat the interaction with the dipole moment on atom j

$$\mu_i = \alpha_i [E_i^0 - \sum_i T_{ij} \mu_j] \tag{3}$$

where E_i^0 is the electrostatic field due to the static charges. In the fluctuating charge, or electronegativity equalization model an atomic dipole is not added but, rather, the partial atomic charges in a molecule are allowed to redistribute to yield equivalent electronegativity on each atom. This leads to a change in the overall molecular dipole moment rather than of the individual atom dipoles as occurs with the induced dipole model. In this approach, each atom type is assigned electronegativity, χ, and hardness, J, parameters where the latter is related to the ability of an atom to transfer charge. Charges are then optimized to minimize the overall electrostatic energy, E_{elec}. with respect to the charges, q_i, via:

$$E_{\text{elec}} = \sum_i \left(\chi_i q_i + \frac{1}{2} J_i q_i^2 \right). \tag{4}$$

Once convergence of the charges has been achieved, the total electrostatic energy is calculated via the Coulombic term in eq. (1). Another alternative is the classical Drude oscillator,[68,69] also referred to as the Shell model. In this model an additional "Drude" particle is attached to the nucleus of each atom, with a charge, $q_{i,D}$, assigned to the Drude particle and the atomic charge, q_i, set to $q_{i,0} + q_{i,D}$, where $q_{i,0}$ is the static partial atomic charge. The atomic polarizability, α, is determined via $\alpha = {q_{i,D}}^2/k$, where k is the force constant of a harmonic spring between the nucleus and the associated Drude particle. Polarization of the system is attained by allowing the positions of the Drude particles to relax in the external electrostatic field, with the electrostatic energy again being obtained via the Coulombic term in eq. (1) where all the atomic and Drude particle charges are taken into account. The Drude oscillator method has only been used in a few studies thus far.[50,70–72] In all of the approaches used to explicitly treat electronic polarization, the polarizability may be solved iteratively, analytically, or, in the case of MD simulations via extended Lagrangian methods.[12,59] Extended Lagrangian methods treat the polarizability as a dynamic variable in the simulation. These methods are important for the inclusion of polarizability in biomolecular force fields as they offer the necessary computational efficiency to perform simulations on large systems.

To date, the majority of work on polarizable force fields has involved water, where polarizable water models have been shown to accurately treat both the gas and condensed phase properties. Examples include models by a number of workers.[50,55–58,60,73–75] In many of these models there was an increased ability to more accurately treat changes in properties, such as the density as a function of temperature, compared to additive water models. Other examples where the inclusion of electronic polarization has lead to improvements in the modeling of molecular interactions includes the solvation of ions,[54,61,70,76] ion-pair interactions in micellar systems,[77] condensed phase properties of a variety of small molecules,[60,65,78–81] cation–π interactions,[82] and in interfacial systems.[83] However, although application of polarization to small molecules has made progress, developments in the area of biomacromolecular force fields have been limited. The first reported study of macromolecular simulations involved proteins in the absence of solvent, with the duration of the simulations being only 2 ps.[84] More recently, simulations of several small proteins using a fully polarizable force field, including solvent and for time durations into the ns range[85] and a simulation of DNA in solution, including counterions,[86] have been reported. Thus, progress is being made towards the development of polarizable force fields for biomolecules.

Although the reasons for the lack of success of polarizable models in macromolecular systems are not totally clear, one possibility is the inability to directly transfer gas phase molecular polarizabilities to the condensed phase.[50] Such an effect is associated with the Pauli exclusion principle such that the flexibility of the electron cloud to distort due to induction by the environment is hindered by the presence of adjacent molecules in the condensed phase.[87] This would cause a tendency towards overpolarization in condensed phase environments when gas phase polarizabilities are applied directly. Further studies are required to better understand this phenomenon.

Despite the problem mentioned in the previous paragraph, it may be anticipated that in the near future successful application of multiple polarizable models to biomolecular simulations will occur. Only once that occurs and highly refined force fields that include polarizability have been developed, will it be possible to judge the improved accuracy and utility of such models over current additive models. However, it is likely that for many applications significant gains in accuracy will not be obtained using polarizable models, ensuring the continued utility of the additive force fields currently in use. Furthermore, in situations where sampling is more important than higher accuracy, the lowered computational costs of additive models are advantageous.

Combining Rules

One of the important considerations that limits the mixing of parameters from different force fields are the combining rules used for the LJ interactions. Combining rules are used to take the LJ parameters for individual atoms (e.g., well depth, ε_i, and minimum radius, $R_{\min,i}$, in eq. (1) and combine them to yield the atom i-atom j LJ interactions for each specific atomic pair. For example, in CHARMM and AMBER the "combined" ε_{ij} values are obtained via the geometric mean, $\varepsilon_{ij} = \mathrm{sqrt}(\varepsilon_i * \varepsilon_j)$, and $R_{\min,\,ij}$ via the arithmetic mean, $R_{\min,\,ij} = (R_{\min,\,i} + R_{\min,\,j})/2$, referred to as the Lorentz–Berthelodt rules. Alternatively, OPLS applies the geometric mean for both the well depths and radii[88] and other methods have been discussed.[33] In cases where the combining rules for two force fields differ, transferring parameters between those force fields is typically not recommended. If such a transfer is performed, testing of the parameters with the new combining rules via, for example, pure solvent simulations, is recommended. Also, the user should be aware of the use of $R_{\min}$ vs. σ for the radius, where $R_{\min}$ represents the radius at which the LJ function is a minimum, as used in AMBER and CHARMM, vs. where the LJ energy is zero in the case of σ, as in OPLS.

1,4 Interactions

In empirical force fields the 1,2 (i.e., nonbond interactions between bonded pairs of atoms) and 1,3 (i.e., nonbond interactions between atoms separated by two covalent bonds) nonbonded interactions are neglected, such that the forces between 1,2 and 1,3 atom pairs are dictated by the internal parameters, the bonds and valence angles. 1,4 Nonbonded interactions, involving atoms separated by three covalent bonds, do have nonbonded contributions as well as their spatial relationship being influenced by the internal terms, including the dihedral term. However, different biomolecular force fields treat the 1,4 nonbond interactions differently. This is important as the 1,4 interactions contribute to the conformational energies, influencing both the relative energies of minima and the barriers betweens those minima. The all-atom CHARMM biomolecular force fields do not scale the 1,4 terms (scale factor = 1). In the original AMBER force fields a 1,4 scale factor 0.5 was used.[35] Due to difference in charge distributions in OPLS, a 1,4 scale factor of 0.83 (1/1.2) was used;[88] this value has subsequently been adopted in the new AMBER protein and nucleic acid all atom force fields.[17] Thus, the use of different 1,4 scaling factors further limits the ability to transfer parameters between different force fields.

Lone Pairs

In early force fields lone pairs were included to improve the treatment of interactions between molecules, as in the ST2 water model.[89] However, it has been shown that hydrogen bonding can be adequately treated based on models that only include charges on atomic centers.[34] In certain cases, such as the ability to reproduce QM data for in-plane vs. out-of-plane interactions between pyridine and water[90] the inclusion of lone pairs improves accuracy. Indeed, the recently developed TIP5P water model contains lone pairs.[91,92] In general, it may be assumed that the inclusion of any additional centers to a force field will improve the accuracy due to the increased number of parameters available. However, such additional centers could lead to complications in parameter optimization due to the parameter correlation problem.[36] For example, the presence of lone pairs may further complicate the determination of partial atomic charges via fitting to QM electrostatic potentials (ESP), which is already problematic as emphasized by the need to include restraints (RESP) when using this approach.[93]

All-Atom vs. United Atom Force Fields

Biomolecular force fields can explicitly treat all the atoms in the molecule, referred to as an all-atom force field. Alternatively, the hydrogen atoms can be neglected to different extents, with the nonbond parameters of the atom to which the hydrogen was attached adjusted accordingly. These are referred to as extended or united atom force fields. In most cases, united atom force fields explicitly include polar hydrogens, facilitating the treatment of hydrogen bonding. Although many of the aspects of intermolecular interactions can be accounted for in united atom models, certain interactions will be poorly treated, an example being aromatic-aromatic interactions.[94,95] Currently, both all-atom and united-atom force fields are used for biomolecular simulations, as discussed in more detail below.

Treatment of Solvation

Accurate treatment of the condensed aqueous environment is obviously an essential aspect of a force field being used for biomolecular simulations. Such treatment may be performed using explicit or implicit models, with the former being a more microscopically complete method while the latter having the advantage of savings in computer time as well as directly yielding free energies of solvation. Explicit water models used in biomolecular simulations include the

TIP3P, TIP4P,[96] SPC, extended SPC/E,[97] and F3C[98] models. All of these models yield satisfactory agreement for bulk water at ambient temperatures. TIP3P is probably the most commonly used model; limitations include underestimation of the height of the second or tetrahedral peak in the O—O radial distribution function and a diffusion constant significantly larger than experiment.[99] On the other hand, the model does treat energetics satisfactorily and because the majority of biomolecule–water interactions involve first or second shell hydration, the lack of long-range structure is often not problematic. The SPC models are similar to the TIP3P, though, by using a tetrahedral geometry (i.e., H–O–H angle = 109.47°) have increased structure as evidenced by a more-defined tetrahedral peak in the O—O radial distribution function. 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. Such an overestimation may perturb the energetic balance of solvent–solvent, solute–solvent, and solute–solute interactions and, therefore, must be considered when using this model in biomolecular simulations. The TIP4P model, which includes an additional particle alone the H—O—H bisector, overcomes many of the limitations listed above at the expense of increased computational costs due to the additional particle in the model. In addition, recent advances in additive water models have been made,[91,100] although their use in biomolecular simulations has not yet been widely validated. Finally, when selecting a water model to use for a particular study, the most important consideration is compatibility with the biomolecular force field being used. This is 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[101,102] have made significant advances in recent years, as recently reviewed by Feig and Brooks.[103] Such models offer significant computational savings while yielding an accurate treatment of solvation. These approaches are typically used in studies were extensive sampling of conformational space is required, such as in protein folding. However, these models can fail when highly specific water–biomolecule interactions are important. Early implicit solvation models included simple treatments such as distance-dependent dielectric constants (r-dielectric) and atomic solvent accessibility based free energy contributions to solvation.[104] A significant improvement in the treatment of solvation is the use of the Poisson–Boltzmann (PB) model, where contributions from solvent polarization along with the asymmetric shapes of biological molecules are taken into account.[105] Improvements in the accuracy of PB methods have been obtained by optimizing atomic radii to reproduce experimental free energies of solvation of model compounds representative of biomolecules.[106,107] Although PB methods can yield free energies of solvation, they are relatively computationally expensive making them of limited value for molecular dynamics (MD) simulations. An effective alternative are generalized-Born (GB)-based solvation models.[108] A variety of GB models have been developed, having computational speeds significantly enhanced over PB or explicit solvent treatments, while yielding free energies of solvation with accuracy comparable to the PB methods.[109–113] In addition, both the PB or GB methods can be combined with free energy solvent accessibility (SA) terms that account for the hydrophobic effect,[114,115] referred to as PB/SA or GB/SA approaches. Recently, a GB-based implicit solvent model was introduced that includes a more rigorous treatment of van der Waals dispersion contributions beyond solvent accessibility related terms.[116] Other implicit models that have been used in biomolecular simulations include the Langevin Dipoles Model[117] and the EEF1 model.[118]

A useful application of PB/SA and GB/SA methods is to first perform MD simulations of biomolecules 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).[119] 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 and small molecule–biomolecule complexes.[120,121] This type of approach should also be useful for the estimation of ligand–protein binding,[122] at a computationally reasonable cost as required for testing of large numbers of drug candidates.

Treatment of Long-Range Interactions

The most important forces dictating the properties of biomolecules are the nonbonded interactions. However, the determination of these interactions represents the most computationally expensive part of empirical force field calculations. Accordingly, approaches were developed to truncate both the electrostatic and LJ atom–atom interactions at selected distances, thereby limiting the number of pair-wise interactions that had to be computed.[14] In early force fields, truncation distances in the range of 8 to 9 Å were typical, with the force field being optimized using those truncation or cutoff distances. In general, it is most desirable to use a given force field with the truncation distance for which it was developed. An example of the importance of this is the TIP3P water model, which was developed with the electrostatic interactions truncated at 8.5 Å with a cubic switching function used to smooth the potential energy to zero over the final 1 Å.[96] Later studies using the TIP3P model with longer truncation distances or with long-range interactions treated via Ewald methods (see below) showed a significant increase in the diffusion constant of the model above the experimental value.[99] However, more accurate treatment of biomolecular systems is generally expected via inclusion of more or, ideally, all long-range interactions. This may be obtained simply by increasing the truncation distance as allowed by increases in computer power with, importantly, use of proper functions that gradually smooth the nonbond interaction energies[14] and forces[123] to zero. Alternatively, and ultimately more desirable, are methods that allow for all long-range interactions to be taken into account. With respect to the LJ term, methods to account for the truncated LJ contribution have been developed that typically treat the region beyond the truncation distance as being homogenous (i.e., an "average" over all LJ atom types in the system).[124,125] It is often assumed that the LJ contribution at longer distances is insignificant due to the r^6 distance dependence; however, it should be emphasized that all the truncated contributions are favorable,

such that summation over those contributions becomes significant, especially in systems dominated by aliphatic groups, as in lipids. For the electrostatic interactions various multipole methods, where, beyond the truncation distance, interactions are treated as atom–multipole interactions,[126–128] were developed and shown to be useful. These have, to a large extent, been replaced by the use of the Ewald summation method,[4] that takes advantage of crystal symmetry combined with reciprocal space to treat the long-range electrostatic interactions (i.e., periodic boundaries are, typically, of cubic or orthorhombic symmetries). Of particular impact has been the particle mesh Ewald (PME) method[129] that has been implemented in a number of the widely used simulation packages; a number of other Ewald-based implementations also exist.[130–132] When using Ewald methods it is important to use a periodic system of adequate size to avoid possible artifacts associated with the biomolecule "feeling" an electrostatic potential identical to itself in the periodic environment (e.g., electrostatic interactions between the primary biomolecule and its images) that may damp the dynamics of the system.[133,134] To avoid such artifacts, use of a minimum of a 7 Å water layer beyond the macromolecule in each direction of the periodic system is recommended.[135] In many cases the macromolecular size disallows use of a periodic system; in such cases alternatives based on the stochastic boundary approach may be used.[136,137] In this approach a spherical system is created that is surrounded by a potential that maintains the density of the system. This potential includes a reaction field that accounts for particles in the region beyond the sphere that have been omitted and, in certain cases, will apply Langevin dynamics to waters at the edge of the sphere to supplement the reaction field. These methods can be used with multipole approaches to treat long-range electrostatic interactions. Recent developments in stochastic boundaries allow for adjusting the sphere size to produce a constant pressure scenerio as well as reaction fields that represent heterogeneous atom distributions associated with, for example, the region of the protein beyond the sphere that was deleted.[6,7] Recent applications of these methods combined with potential of mean force calculations have yielded novel insights into ion channels[138] and base flipping in DNA.[139] In the end, when setting up a simulation system the balance between accuracy and computational feasibility must be taken into account. Periodic boundary conditions may be considered the most rigorous approach and, therefore, should be considered first, followed by stochastic boundary approaches. In all cases, it is recommended that the approach selected be tested on model biomolecular systems for which adequate experimental data exists to validate the approach. Finally, although most of the current force fields were developed using specific atom truncation methods, it is considered more appropriate to treat all the long-range contributions to both the LJ and electrostatic terms.

Biomolecular Force Fields

To understand the applicability of a given force field it is important to be aware of the approaches used in its development. Accordingly, an overview of the assumptions in and the methodologies used for the optimization of current biomolecular force fields will be presented. This will be followed by a survey of the force fields used for the different classes of biomolecules, with emphasis placed on the range of applicability of the various force fields. With respect to protein force fields, a recent review by Ponder and Case nicely supplements the present review.[52]

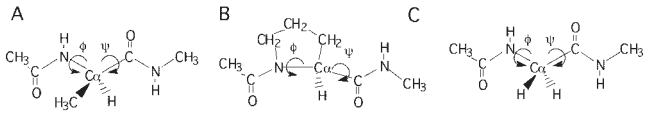

Figure 1. (A) Alanine, (B) proline, and (C) glycine dipeptides used for parameter development of the protein backbone. Shown in the figures are the ϕ,ψ dihedral angles that define the Ramachandran map.[169]

Force Field Optimization

As stated above, a potential energy function is not a force field until the parameters are available that allow for the energies and forces of the selected molecules to be calculated. It is the approaches used to obtain the parameters that ultimately dictate the applicability and quality of the force field. During such development a very important step is the selection of the target data to be used as the basis for the parameter optimization. For a variety of force fields this target data is comprised of QM results for model compounds representative of the biomolecules of interest. For example, the alanine dipeptide is the quintessential model for optimization of the protein backbone parameters, often being supplemented with data from the glycine and proline dipeptides (Fig. 1) as these represent the amino acids in which the covalent structure of the backbone differs from the remaining 18 natural amino acids. QM data can be obtained on the geometries, allowing for optimization of the bond and valence angle equilibrium constants and the dihedral multiplicity and phase, and on the vibrational spectra, including assignments of the normal modes,[140] allowing for adjustment of the force constants. Conformational energies as a function of rotation about selected bonds (e.g., ϕ,ψ in Fig. 1) from QM calculations are additional target data for the optimization of the dihedral parameters.[141,142] Although QM data is beguiling due to the significant amount of information that can readily be extracted and used as target data and, with increases in computer power, can be applied to larger systems, including small peptides,[143] nucleosides and nucleotides,[144,145] "exact" reproduction of QM data by a force field designed for condensed phase calculations may be inappropriate due to the lack of condensed phase contributions in the QM calculations as well as issues associated with the level of theory. A good example occurs with the geometry of the peptide bond (Table 1).[146] In the gas phase the peptide bond, C—N, is relatively long based on both experimental and QM data; however, upon moving to the condensed phase there is a significant decrease in the bond length, again seen both experimentally and in QM supramolecular calculations where water or formamide molecules are included to mimic the condensed phase environment. The opposite effect occurs with the C═O bond and analogous variations occur with the C_m—C—N and C_m—C═O angles. Thus, use of the gas phase QM data directly would yield geometries not appropriate for the condensed phase.

Table 1. Comparison of Peptide Bond Geometries from QM and Experimental Methods.

	Experimental			MP2/6-31 G(d)[b]		
	Gas[c]	Crystal[d]	Survey[e]	Gas	$3H_2O$	H_2O,2FM
Bonds						
C_m—C	1.520 (5)	1.515 (3)	1.52 (1)	1.514	1.510	1.512
C—N	1.386 (4)	1.325 (3)	1.33 (1)	1.365	1.339	1.337
N—C_m	1.469 (6)	1.454 (3)	1.45 (2)	1.448	1.454	1.454
C═O	1.225 (3)	1.246 (2)	1.23 (1)	1.232	1.255	1.254
Angles						
C_m—C—N	114.1 (15)	116.3 (6)	116 (2)	115.3	117.1	116.6
O═C—N	121.8 (4)	121.7 (6)	123 (1)	123.1	122.1	122.6
C_m—C═O	124.1	121.9 (6)	121 (4)	121.6	120.9	120.9
C—N—C_m	119.7 (8)	121.3 (6)	122 (1)	122.1	121.1	121.3

Bonds and angles in Å and degrees, respectively. C_m indicates a terminal methyl carbon. Values in parenthesis represent the standard deviation in the final digit(s).

[b]From ref. 324, $3H_2O$ indicates two water molecules hydrogen bonding to the carbonyl oxygen and one water molecule hydrogen bonding to the amide proton; H_2O,2FM indicates one water molecule and one formamide hydrogen bonding to the carbonyl oxygen and one formamide hydrogen bonding to the amide proton; see original reference for the exact geometries.

[c]Gas phase electron diffraction data from ref. 325.

[d]Crystal values are from ref. 326 for the 0.9 occupancy structure.

[e]Survey of the Cambridge Crystal Data Bank from ref. 161.

Similar effects also occur with conformational energies. An example is the rotatable bonds in the phosphodiester backbone of DNA and RNA. Extensive QM calculations on model compounds representative of the backbone have yielded a variety of energetic target data[147–149] for optimization of the associated dihedral parameters. However, accurate reproduction of that data leads to systematic deviations in the sampling of dihedrals in the phosphodiester backbone from survey data from the nucleic acid database (NDB).[150] An example for the γ dihedral in DNA is shown in Figure 2.[48] QM and empirical potential energies as a function of γ are shown in Figure 2A for the model compound included in the figure. The initial energy surface (open squares) was based on parameters (set 1) optimized to reproduce the QM surface in the region of 60°, the region of γ that is sampled in duplex DNA and RNA crystal structures. However, when this parameter set was applied in an MD simulation, the resulting γ probability distribution (Fig. 2B) was too narrow compared to NDB survey data. Subsequent alterations of the parameters to systematically "soften" the energy surface leading to poorer reproduction of the QM energy surface (closed squares and diamonds) led to improved agreement of the MD distribution with the NDB data. This empirical or knowledge-based optimization approach was used for the remainder of the dihedrals in the CHARMM27 nucleic acid force field and has been used elsewhere for the optimization of nucleic acid and protein force fields (see below). With respect to the peptide backbone, a novel extension of the potential energy function for the treatment of the ϕ,ψ dihedrals has allowed for the relationship of QM conformational energy surfaces to properties obtained from simulations of proteins in crystal environments to be tested more rigorously,[29] further indicating limitations in the use of QM data directly in certain cases. However, it should be emphasized that the capability to carefully examine geometric, vibrational, and conformational properties, allowing for quantification of condensed phase contributions, is limited to systems for which extensive experimental data is available, such as nucleic acids and proteins. In such scenarios, careful parameter optimization taking experimental data into account can yield improved accuracy in the force field. However, for many systems such experimental data is

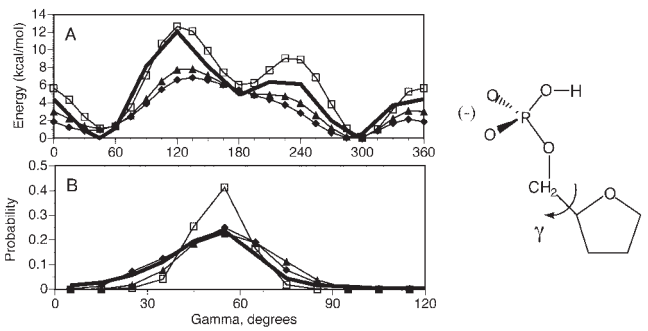

Figure 2. Potential energies (A) and probability distributions (B) as a function of the γ dihedral. The potential energy surfaces (A) were obtained using the presented compound at the QM HF/6-31+G* (bold line) level of theory and for three empirical parameter sets designated 1 (open squares), 2 (triangles), and 3 (diamonds). Probability distributions are from the NDB survey (bold line) from crystal simulations of the CGATCGATCG B form decamer using the same three empirical parameter sets 1 (open squares), 2 (triangles), and 3 (diamonds). Note the change in the X-axis upon going from A (0 to 360°) to B (0 to 120°). See ref. 48 for methodological details.

not available (e.g., drug-like molecules), such that QM results represent the best target data available for parameter optimization.

Due to the dominant role of the nonbond or external parameters in dictating thermodynamic properties and their role in biomolecular structure and interactions, proper optimization is essential for any successful biomolecular force field. A large number of studies have gone into the determination of the electrostatic parameters, that is, the partial atomic charges. For biomolecular force fields the most common charge determination methods are the QM electrostatic potential (ESP) and supramolecular approaches. Other variations include bond charge increments[151,152] and electronegativity equilization methods.[153] ESP-based methods are based on the optimization of charges to reproduce a QM determined ESP mapped onto a grid surrounding a model compound. Such methods are convenient, and a number of charge fitting methods based on this approach have been developed.[154–157] They allow for charges to readily be determined for any molecule accessible to the appropriate QM level of theory. However, the ability to unambiguously fit charges to an ESP is not trivial,[158] and charges on "buried" atoms tend to be underdetermined, requiring the use of restraints during fitting,[93] a method referred to as Restrained ESP (RESP) fitting. In addition, because the charges are based on a gas phase QM wave function, they may not necessarily be consistent with the condensed phase, although recent developments are addressing this limitation.[159] Also, consideration of multiple conformations must be taken into account.[160] An alternative is the supramolecular approach where the charges are optimized to reproduce QM determined interaction energies and geometries of the model compound with, typically, individual water molecules, although model compound dimers are often used.[88,161] This approach is somewhat tedious in that the supramolecular complexes used must be created individually and the QM calculations performed followed by the empirical calculations. However, in this approach the interacting pair in the QM calculation leads to local electronic polarization, which is then implicity included in the obtained charges; a result important for additive force fields where implicit inclusion of electronic polarizability is required (see above). It should be noted that in CHARMM and OPLS, whose charges are obtained using the supramolecule approach, functional groups are assigned integer charges, allowing those charges to be transferred between molecules, facilitating the assignment of charges to novel molecules. In both the ESP and supramolecular approaches the QM level of theory of choice for additive force fields has been HF/6-31G* as it is known to overestimate dipole moments[160] and interaction energies. Such overestimations are again desirable for additive force fields, as they lead to partial charge distributions that include the implicit polarization required for condensed phase simulations. Also, it should be reiterated that charges from both the supramolecular and ESP approaches are conformation dependent, requiring care to select the appropriate conformation(s) when performing the charge determination.

Proper optimization of the vdW or LJ parameters is one of the most important aspects in the development of a force field for condensed phase simulations. Early efforts by Jorgensen developed the use of condensed phase simulations, typically neat liquids, as the basis for optimization of the LJ parameters,[37,162] Typically, once partial atomic charges were assigned via the supramolecular approach, the LJ parameters for a model compound were adjusted to reproduce the experimentally determined heat of vaporization and density as well as isocompressibilities and heat capacities when available. Similarly, heats or free energies of aqueous solvation or heats of sublimation and lattice geometries[46,163] can be used as the target data for the LJ optimization. Although such methods are effective, the parameter correlation problem allows for LJ parameters for different atoms in a molecule (e.g., H and C in ethane) to compensate for each other such that it is difficult to accurately determine the "correct" LJ parameters of a molecule based on reproduction of condensed phase properties alone.[36] To overcome this problem a method has been developed that determines the relative value of the LJ parameters based on high level QM data[164] and the absolute values based on the reproduction of experimental data.[43,45] This approach is again tedious, as it requires supramolecular interactions involving rare gases; however, once satisfactory LJ parameters are optimized for atoms in a class of functional groups they can often be directly transferred to other molecules with those functional groups without further optimization.

Finally, the correlation among parameters in the potential energy function should be emphasized. First, the LJ parameters and partial atomic charges are highly correlated, such that LJ parameters determined for a given set of charges are typically not appropriate for charges determined via another methodology. Second, the internal parameters are dependent on the nonbond parameters. For example, the energy surface for rotation about a bond will typically be dominated by the dihedral term, but will also contain contributions from the electrostatic and LJ terms.[165] Thus, changing, for example, the partial atomic charges will lead to changes in the energy surface, requiring readjustment of the dihedral parameters. Such correlations are another factors leading to the inability to combine parameters from different force fields and still maintain the proper balance of the intra- and intermolecular forces. In addition, the importance of using the correct water model with a given force field must be reiterated, as the nonbonded parameters in a force field are optimized to be compatible with a specific water model. Accordingly, application of an alternate water model with the force field may yield spurious results.

Protein Force Fields

The first successful MD simulation of a biological macromolecule[166] was performed on a protein and protein simulation studies continue to dominate the field. Current protein simulations use both united-atom and all-atom force fields, although the majority of protein MD simulation studies, excluding protein-folding studies, are performed using all-atom protein models, including the OPLS/AA,[88] CHARMM22,[161] and AMBER (PARM99)[17] force fields.

Parameters for all three force fields were extensively optimized with particular emphasis on the treatment of proteins. With OPLS and CHARMM22 the partial atomic charges were 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. All three force fields used condensed phase simulations on an extensive set of model compounds to determine the LJ parameters. CHARMM22 and AMBER were primarily optimized based on the

TIP3P water model while OPLS was developed to work with the TIP3P, TIP4P, and SPC models. It should be noted that the similarities of the water dimer interaction energy of these three water models should allow the TIP4P and SPC models to be used with CHARMM22 and AMBER, although rigorous tests have not been performed. Thus, all three force fields model intermolecular interactions well, although differences in the local charge distributions have been pointed out,[52] which may lead to differences in the balance of the local interactions (e.g., relative hydrogen bonding strength at the peptide bond NH vs. CO). It should be emphasized that such differences are important,[45] as the goal of molecular simulations is to elucidate atomic details of the properties of biomolecules. Accordingly, when analyzing nonbond interaction results from MD simulations the method used in the optimization of the intermolecular parameters should be considered with respect to its potential influence on the results.

Intramolecular parameters for CHARMM22 and AMBER were both derived based on reproduction of a variety of experimental and QM data for small model compounds. This included emphasis on the reproduction of vibrational spectra for optimization of the force constants. With OPLS, the internal portion of the force field was originally taken from AMBER PARM94[17] followed by optimization of selected torsion parameters based on QM data, yielding OPLS/AA.[40] OPLS/AA has subsequently been updated via reoptimization of selected torsions using higher level QM target data.[167] Similar reoptimization of the torsional terms has occurred with the other force fields, as discussed in the following paragraph. Overall, accurate treatment of the internal portion of the biomolecular force fields via reproduction of vibrational and conformational energy data ensures that the intramolecular distortions the molecules undergo during MD simulations will be representative of the experimental regimen. Supporting the quality of the three force fields in attaining such a representation is a recent comparison of these three force fields in MD simulations of three proteins, showing them to reproduce the experimental structures in a similar fashion.[168]

An important issue in protein force fields is the treatment of the conformational energies associated with ϕ,ψ (i.e., the Ramachandran map,[169] Fig. 1). The relative energies as a function ϕ,ψ will have a significant influence on the sampling of conformational space by the protein backbone in MD simulations. Typically, the alanine dipeptide (Fig. 1) is the model compound used to study ϕ,ψ conformational energies. This, and related compounds, are small enough to allow for QM calculations at high levels of theory[170–172] the results of which may then be used as target data for force field optimization. A comprehensive study of the ability of a variety of empirical force fields to reproduce QM data on the alanine dipeptide and the alanine tetrapeptide has been reported.[142] In that work it was concluded that certain force fields were "better" based on their ability to reproduce the QM target data. However, other studies have shown that exact reproduction of QM data may, in fact, leads to less accurate agreement with respect to experimentally determined properties on proteins.[161]

Recently, the relationship between the reproduction of gas phase QM data on the alanine dipeptide by a force field with the sampling of ϕ,ψ in MD simulations was investigated in greater detail exploiting an extension of the potential energy function.[29,30] Using the sinusoidal dihedral function common to current potential energy functions [eq. (1)] limits the ability of protein force fields to accurately reproduce QM energetic data for the entire ϕ,ψ surface. To overcome this, extensions of the potential energy function were tested, including ϕ,ψ crossterms and a 2D dihedral energy grid correction map (CMAP) approach.[29,30] Using the CMAP approach the CHARMM22 protein force field was modified to reproduce the LMP2/cc-pVQZ//MP2/6-31G* gas phase energy surface of the alanine dipeptide. This force field was then applied to MD simulations of proteins in their crystal environments. From those simulations systematic differences in ϕ and ψ between calculated and experimental crystal structures were observed. Motivated by these systematic differences, empirical adjustments to the alanine dipeptide conformational energy surface were undertaken. The resulting force field significantly improves the sampling of ϕ,ψ in MD simulations of selected proteins as judged by the reproduction of survey data from the PDB,[173,174] as shown in Figure 3. In addition, the modified CHARMM22 force field also better reproduces the sampling of ϕ,ψ conformational space of the alanine dipeptide in solution (Fig. 3, lower panels) compared to results from previously published QM/MM calculations.[175,176] Subsequently, additional empirical optimization of the alanine dipeptide conformational energies, empirical optimization of the glycine dipeptide surface and inclusion of a proline dipeptide surface yielded a CHARMM22/CMAP force field included in version 31 of CHARMM.[30] Although additional tests verifying the accuracy of this approach are ongoing (Rich Pastor, personal communication; Peter Steinback, personal communcation; Matthias Buck, personal communication), the inclusion of the 2D

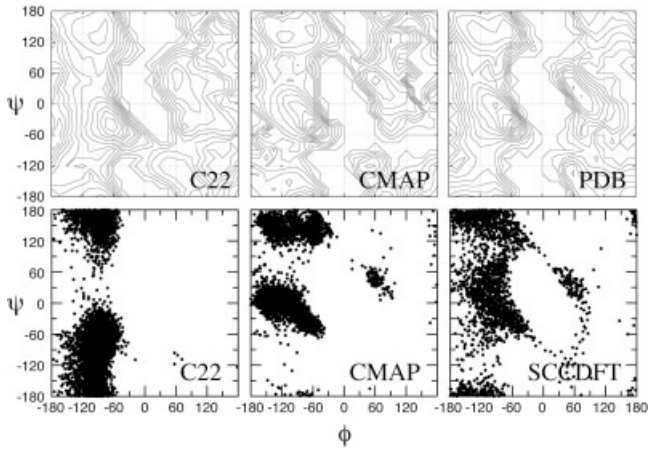

Figure 3. ϕ,ψ PMFs based on MD simulations using the CHARMM22 and CHARMM22 grid-corrected (CMAP) empirical force fields and from a survey of the PDB (upper frames) and ϕ, ψ distributions from MD simulations of the alanine dipeptide (Ace-Ala-Nme, lower frames) in solution using the CHARMM22[161] and CHARMM22 grid-corrected empirical force fields and previously published data from a QM/MM model (SCCDFTB). 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.[199] See ref. 29 for more details. Reproduced with permission from J Am Chem Soc 2004, 126, 698–699. Copyright 2004 Am Chem Soc.

dihedral energy grid correction appears to significantly improve the treatment of ϕ,ψ sampling in empirical force fields by allowing accurate treatment of both low and high energy regions of ϕ,ψ space. It is anticipated that this approach will have general applicability in empirical force fields for proteins.

The importance of accurate treatment of ϕ,ψ conformational properties is emphasized by the observation of the presence of π helical structures in MD simulations of peptides using a variety of force fields.[177] In these studies, which typically involved helical peptides, sampling of π helical conformations was observed. Careful analysis of the relative energies of the α_R vs. π helical conformations in a variety of force fields revealed the energy difference to be smaller than that predicted by high level QM calculations. This suggested that, because the energy of the π conformation was not sufficiently high compared to the α-helical conformation, additional sampling of π was occurring. Verification of this was performed by applying the CMAP approach where, by reproducing the π to α-helical QM energy difference in a modified CHARMM22 force field, sampling of the π-helical conformation was significantly diminished. Although these results do not eliminate the possibility of π-helical conformations in proteins,[178] they emphasize the influence a force field can have on the results of MD simulations, and that when novel results are obtained from empirical force field studies, underlying influences associated with the force field being applied should be considered.

The improvements attained with the 2D dihedral CMAP correction emphasize the need for continuous optimization of empirical force fields as technical developments allow for more rigorous comparison of simulation results with experimental data. A good example is a recent study in which free energies of solvation of model compounds representative of protein side chains were calculated for the AMBER, CHARMM, and OPLS-AA force fields and compared with experiment,[49] similar studies have been reported elsewhere.[179–181] In the study all three force fields were shown to satisfactorily reproduce the experimental data, with OPLS showing better overall agreement, while CHARMM and AMBER were similar. However, in certain cases all of the force fields performed poorly. Such results, while somewhat limited because of the use of atom truncation methods vs. the more commonly used PME methods for the treatment of long-range electrostatics, indicate that improvements in the nonbond parameters of all the force fields are possible.

Multiple adjustments of the original AMBER Cornell et al force field,[17] referred to as PARM94, have occurred. The first significant adjustment, yielding PARM96, included optimization of the dihedral parameters associated with ϕ,ψ to yield better agreement with the alanine dipeptide results of Beachy et al.[142] However, in a subsequent version, PARM98, the ϕ,ψ related torsions were set back to the PARM94 values. Additional adjustments of the ϕ,ψ torsions were again undertaken targeting QM data for both the alanine dipeptide and tetraapeptide,[142] yielding PARM99. More recently, additional modifications have been performed motivated by a tendency for the AMBER force field to favor α-helical conformations. These include modifications to the ϕ,ψ related dihedral parameters and to the charge distribution for the entire protein force field. In two studies, only the dihedral parameters were changed to better treat conformational sampling in peptide simulations,[182,183] with the dihedral force constants being set to zero in one case.[182] In a third study, the partial atomic charges for AMBER were redetermined via RESP fits to B3LYP/cc-pVTZ//HF/6-31G* QM data.[184] 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. The use of a dielectric constant of 4 was designed to mimic that of the protein interior. The resulting model yielded improved reproduction of conformational properties for selected peptides using a GB model[185] of solvation. However, it should be noted that changing the charge distribution of a protein force field should be accompanied by reevaluation of the internal portion of the force field, as the electrostatic term effects geometries, vibrational spectra, and conformational energies to varying degrees. In addition, care must be taken with the use of individual charge sets for the main chain atoms of the different amino acids (i.e., different partial atomic charges on the mainchain atoms for the different sidechains) as the difference could lead to poor ϕ,ψ distributions, although the authors of the Duan et al. study suggest that it may lead to improved sequence-dependent properties.[184] Overall, these changes in the AMBER force field represent positive steps, potentially leading to improved models. Clearly, additional tests on these models via MD simulations of larger proteins using explicit representations of the condensed phase environment are required for their further validation. Also, care must be taken to avoid the creation of a number of AMBER "variant" force fields, which may lead to problems in comparing results from different studies as well as make it difficult to perform further enhancements of the AMBER force field in a coherent fashion.

Extended or united atom force fields dominated the early protein force fields. Examples include OPLS/UA,[88] the early AMBER force fields,[35] GROMOS87 and 96[186] and CHARMM PARAM19.[187] The GROMOS united atom force field[186] is still widely used in MD simulations that include explicit solvent representations. Recent enhancements in GROMOS96 have included condensed phase tests[188] and more careful optimization of LJ parameters to reproduce experimental condensed phase properties.[189] Of the other united atom force fields, their use is now primarily for simulations on long time scales via the use of implicit solvent models, with the majority of these studies being based on PARAM19. Several continuum solvation models have been developed for use with PARAM19, including EEF1,[118] ACE,[190] several GB models,[113,191,192] and a buried surface area model by Caflisch and coworders.[193] It should be noted that these and other implicit models can and have been used with all-atom force fields. A summary of recent applications of both united and all-atom protein force fields combined with implicit solvent models has been presented,[103] showing the success of this approach in understanding protein structure–function relationships especially in the area of protein folding.

Additional force fields used for protein simulations include MMFF,[194] CEDAR,[195,196] ENCAD,[197,198] and CVFF,[21] among others. A more comprehensive list may be found it the recent review by Ponder and Case.[52] MMFF was designed to treat a wide range of pharmaceutically relevant molecules as well as biomolecules. Although care was taken in the development of the nonbonded portion in the force field, including use of the buffered 14–7 VDW potential, the omission of condensed phase data from

the optimization appears to limit the quality of the model for protein simulations. However, the wide range of molecules that MMFF can treat make it useful for ligand–protein studies (see below), where it may be desirable to restrain portions of the protein to overcome limitations in the protein portion of the force field. A similar situation exists with CVFF. In the end, when selecting a force field it is important to determine that the force field is appropriate for the problem under study and, during the study itself, to evaluate the results to make certain that they are not being unduly biased by the force field.

Knowledge-Based Protein Force Fields

Although not the primary focus of this review, knowledge-based or free energy force fields should be discussed due to their role in studies of protein folding. Knowledge-based force fields are parameterized to directly yield free energies versus force fields defined in eq. (1) that yield potential energies, with the thermodynamic quantities obtain from statistical mechanics.[199] Of the knowledge-based force fields, one of the first was ECEPP by Scheraga and coworkers.[200,201] This force field also represents one of the earliest force fields for any biological macromolecule, and has continued to be enhanced over time.[202,203] More recently, the UNRES potential has been presented,[204] and it too is undergoing continual enhancements.[205,206] In addition, a variety of other knowledge-based potentials are available,[207–211] although it should be emphasized that this is not a comprehensive list (see the review by Russ and Ranganathan for more information[212]). In addition, it is difficult to judge the quality of these different models as many are used primarily by the groups that developed them, although the CASP competitions may be considered one measure of their quality.[213] It should be noted that as the number of proteins for which 3D structures are available increases, the quality and utility of these types of force fields is anticipated to increase.

Nucleic Acid Force Fields

The polyanionic nature of oligonucleotides has made them a challenge for empirical force field-based calculations,[214] requiring more accurate treatment of the balance between the oligonucleotide's conformational energies and their interactions with the aqueous solvent environment. Early simulations of DNA included making the phosphate moieties neutral,[215,216] the inclusion of "solvated" sodiums as counterions and the use of a distance dependent dielectric constant to mimic the solvent environment.[217,218] Numerous attempts to perform simulations of DNA and RNA with an explicit representation of solvent using the early AMBER,[219] CHARMM nucleic acid,[220] or GROMOS[221] force fields found mixed success. Significant progress in DNA simulations occurred in the mid 1990s when several groups performed successful simulations of DNA in solution.[222–224] These successes were initially attributed to the use of the Ewald method,[4] typically in the particle-mesh Ewald (PME) formalism,[129] for the treatment of long-range electrostatic interactions; however, it was subsequently shown that use of adequate lengths for truncation of electrostatic interactions combined with the appropriate smoothing functions[123] led to stable MD simulations of DNA.[225–227] In addition, improvements in second-generation force fields made significant contributions to the ability to perform stable simulations of oligonucleotides.

Second-generation forces fields for nucleic acids included the Cornell et al AMBER (PARM94)[17] and CHARMM all-atom[163] force fields. Both of these models produced stable simulations, as listed in the previous paragraph, but both had systematic problems. With CHARMM22 there was a strong tendency towards A form duplex DNA structures, even in low salt conditions.[228] This problem lead to a full reoptimization of the CHARMM all-atom nucleic acid force field, yielding CHARMM27.[48,229] With AMBER PARM94 problems associated sugar puckering and helical repeat lead to modification of selected torsion parameters,[230] yielding the PARM98/PARM99 force field. In addition to the modified second-generation CHARMM and AMBER nucleic acid force fields, another carefully optimized force field from Bristol-Myers-Squibb (BMS) has been published.[231]

Optimization of the parameters in these three force fields involved some notable differences. AMBER parameters, in general, are based on small molecules from which internal parameters are obtained via reproduction of geometries, vibrational spectra, and conformational energies. These parameters are then applied directly to larger model compounds representative of nucleic acids, with additional adjustments of selected dihedral parameters performed as required when the direct transfer of parameters from small compounds is insufficient (e.g., for the conformational energies of dimethylphosphate). Partial atomic charges were obtained via RESP fitting for the nucleotides in a B-form conformation.[160] These were then applied directly to the nucleic acids, with subsequent additional optimization of torsion parameters performed to yield PARM98/99.[230] CHARMM27 is similarly based on the optimization of the internal parameters to reproduce small molecule geometries, vibrations, and conformational energies, while charges are based on the supramolecule approach discussed above. In addition, final optimization of selected dihedral parameters was based, in part, on the reproduction of conformational energies of larger model compounds representative of nucleic acids, including nucleosides.[144,147–149] LJ parameters for both AMBER and CHARMM27 were transferred from the protein force fields, although the LJ parameters for selected base atoms in CHARMM27 were optimized to reproduce heats of sublimation of bases. The BMS force field is based on AMBER RESP charges, CHARMM-based internal parameters, and CHARMm/Quanta (i.e., commercial CHARMM, Accelrys Inc.) internal parameters for the sugar and phosphodiester backbone.[232] Importantly, all three force fields included additional optimization of selected dihedral parameters based on the reproduction in MD simulations of experimental structural data for DNA and RNA duplexes. With AMBER, the inclusion of such data involved the changes associated with the PARM98 revision, while such target data was used extensively with the CHARMM27 and BMS force fields. In the case of BMS the final optimization was based primarily on the reproduction of structural properties from surveys of the NDB[150] along with proper treatment of the equilibrium between A and B form DNA as a function of water activity.[233,234] With CHARMM27, the final optimization simultaneously targeted survey data from the NDB and the conformational energies of the larger model compounds representative of nucleic acids. The si-

multaneous reproduction of both the survey and model compound data assures that the final dihedral parameters are not inappropriately biased by one type of target data. Despite notable differences in the approach to parameter development, all three force fields yield reasonable properties in simulations of DNA in solution. However, it should be emphasized that the differences in parameter optimization protocols may lead to differences in the atomic details obtained from the MD simulations. Accordingly, atomic detail results obtained from MD simulations with the force fields should be judged with respect to possible biases in the individual force fields.

All three of these force fields have been subjected to a variety of tests on duplex DNA.[135,235] They all yield stable structures of DNA in solution and, importantly, with the exception of the AMBER PARM98/99 model, yield B form structures in high water activity conditions (e.g., low salt) and A form structures in low water activity conditions (e.g., high salt or ethanol). It should be noted that the AMBER PARM94 force field did reproduce the expected structural properties as a function of water activity; in certain cases it may be preferable to use this model over PARM98/99. As presented in detail elsewhere,[135] these force fields reproduce a variety of experimental observables, including changes in helicoidal parameters as a function of sequence. Individually, certain biases are observed in the individual force fields. With CHARMM27, there is a tendency for the minor groove to be wider than experimentally observed values, although an ongoing debate concerning the role of counterions on minor groove width indicates that further studies are required.[236–240] AMBER tends to underestimate helical twist and favor roll over twist, whereas the BMS force fields yields B-DNA solution structures that are possibly to close to the canonical B form of DNA. This bias with BMS may be due to the dominant role of crystal survey results as target data during the optimization of the force field. It should be reiterated that all three force fields work well with canonical DNA.

An area that has not received a significant amount of study is the use of these force fields with noncanonical structures. Considering the significant structural distortions that DNA must undergo to perform its biological function,[241,242] knowledge of the accuracy of empirical force fields in treating noncanonical structures is essential. Some progress towards this has been undertaken in potential of mean force studies of base flipping in DNA, where the CHARMM27 force field as been shown to yield near-quantitative agreement with NMR imino proton exchange experiments[243] and in an MD study where CHARMM was shown to accurately treat an adenine bulge.[244] Clearly, more systematic studies in this area are required.

RNA, which samples a significant number of noncanonical structures, has also been the subject of MD studies. These calculations have been performed with the AMBER and CHARMM22 and CHARMM27 force fields. To date, MD-based studies of both canonical[223,245] and noncanonical[246–249] structures have been reported, although systematic tests and comparisons of the force fields have not yet been performed.

Over the last several years modifications of the AMBER and CHARMM models have been proposed. One study used high level QM calculations on Watson–Crick and Hoogsteen base pairs at the HF/cc-pVTZ(-f)//LMP2/cc-pVTZ(-f) to optimize off-diagonal LJ terms for the base hetereoatoms.[250] This led to improved agreement between the AMBER and CHARMM22 force fields and the presented QM interaction energies, notably improving the overestimation of the WC interaction energies by AMBER. Although leading to improved agreement with the limited target data used in that study, application of such adjustments in MD simulations must be tested to ensure that the balance of the WC interactions with base stacking and base–solvent interactions will not be altered from that in the original implementations of the force fields. In a second study, new internal parameters have been presented for the AMBER force field that better reproduce the out-of-plane geometries of the base amino groups.[251] The optimization was performed based on QM gas phase geometries of Cyt, Ade, and Gua. However, as pointed out by the authors, when those amino groups are hydrogen bonded they become planar due to delocalization of the lone pair on the nitrogen leading to decreased pyramidalization. A similar phenomenon occurs with the peptide bond.[252] Because the change in the pyramidalization is associated with a change in electronic structure due to hydrogen bonding and, in condensed phase simulations the amino groups are always involved in hydrogen bond interactions, it may actually be more appropriate to optimize parameters for the planar bases while taking care to ensure that the out-of-plane wags of the amino groups are properly treated, as was done in both AMBER[17] and CHARMM27[48] to allow for the appropriate out-of-plane distortions that occur in MD simulations. In general, when making ad hoc alterations to a force field it is important to test the changes in a variety of simulation conditions to ensure that they do indeed lead to an overall improvement in the force field rather than just leading to better agreement for a limited set of target data.

Other force fields are available for nucleic acid simulations, although they have not been widely used. These include GROMOS,[253] MMFF, CVFF,[21] and OPLS.[254] In all cases the necessary parameters for the nucleic acids are present; however, it appears that subtle adjustments of the force fields required for accurate simulations of nucleic acids, as discussed above, have yet to be performed for these models. Another alternative is FLEX, implemented in the program JUMNA, which may be considered a knowledge-based force field for nucleic acids.[255,256] JUMNA includes an internal coordinate representation of DNA, treating solvation with a sigmoidal dielectric screening term rather than with an explicit solvent model. Accordingly, JUMNA allows for conformational studies on oligonucleotides that are significantly larger than that currently accessible to all-atom, explicit solvent representations.

Lipid Force Fields

Lipids represent a significant challenge for empirical force fields, while empirical force field calculations offer great potential in understanding the atomic details of lipid structure and dynamics. Both of these points are due to the liquid–crystalline nature of lipid bilayers in biological membranes at physiological temperatures. This fluid-like nature leads to the absence of high-resolution experimental structures of the lipid bilayers, although crystal structures of certain lipids at lower temperatures (i.e., not in the fluid phase) are available.[257] Accordingly, experimental data on biologically relevant lipid bilayers is limited to low-resolution data, such

as density profiles along the bilayer normal obtained from X-ray and neutron scattering, and various NMR observables such as order parameters and T1 relaxation times.[258,259] This leads to the problem of a lack of macromolecular structural target data for careful optimization of lipid force fields while the accuracy requirements of a lipid force field may be considered to be higher then those for other biomolecules, due to the central role of force fields in understanding the atomic details of biological membranes.

Current lipid force fields include both all-atom and united or extended-atom models. Although the majority of protein and nucleic acid simulations are now performed with all-atom models, the relatively large number of aliphatic hydrogens in lipids leads to significant computational gains when a united-atom model is employed. Accordingly, a large number of MD studies with lipids are performed using united atom models. Currently, several united-atom lipid force fields are in use. These include the GROMOS force field, which has been optimized to reproduce condensed phase properties of alkanes.[260] A variant of the GROMOS force field has been presented[261,262] that includes reoptimized LJ parameters based on long chain alkanes and 5-decene and use of the Ryckaert and Bellmans dihedral potential.[263] Similarly, a united atom force field by Berger et al.[264] included optimization of the LJ parameters to reproduce condensed phase properties of pentadecane. This force field has recently been used in a series of simulations of 150 ns or more, showing that long simulations leading to stable bilayer structures are feasible when the appropriate treatment of electrostatics, for example, PME, is performed.[265] The essential role of the treatment of long-range electrostatics via Ewald methods, vs. atom-truncation, for lipid simulations has been discussed.[266] Another united atom force field by Smondyrev and Berkowitz[267] was based on CHARMM22 force field internal terms and charges,[165] supplemented with internal parameters from AMBER, followed by optimization of selected terms based on lipid-related model compounds. Application of this force field to DPPC yielded an area per head group and aliphatic chain order parameters in good agreement with experiment. Thus, there are a variety of united-atom force fields for lipids available, although the GROMOS96 force field appears to be the most commonly used model. Although such a collection of force fields gives users a choice of potentials, it complicates comparison of results from different studies. Clearly, careful comparison of these varied force fields is required to better understand their utility.

The dominant all-atom model in use is that associated with CHARMM, which treats a variety of phospholipids, including mono[268] and polyunsaturated lipids.[269] The history of the lipid parameter development with CHARMM is a good example of how empirical force fields can be continually improved as additional experimental data becomes available, improvements in algorithms are made and, importantly, increases in computer time allow for more rigorous testing of force fields. The original CHARMM force field (CHARMM22) was published in 1996.[165] Optimization of the force field targeting small molecules in the gas phase was performed without atom truncation, while condensed phase tests employed atom truncation. This force field was used in a number of studies, including lipids,[270] a micelle,[271] and protein–lipid complexes,[272] reproducing a variety of experimental observables. However, application of that force field in crystal simulations of glycerolphosphorylcholine and cyclopentylphosphorylcholine yielded densities that were too high.[273] This, combined with limitations in the nucleic acid force field, led to reoptimization in the phosphate parameters, yielding improved agreement with experiment. In addition, it was shown that the heat of vaporization of long chain alkanes was significantly overestimated and, based on high level QM calculations,[274] that the energy of the *gauche* conformation of butane was to high. Application of new LJ parameters for the alkanes[43] combined with lowering the *gauche* energy of butane along with the improved model for the phosphate moiety yielded the CHARMM27 force field that reproduces a variety of experimental observables for a dipalmitoylphophatidylcholine (DPPC) bilayer. An interesting observation during this stage of the optimization was that improvements in the LJ parameters for the aliphatic chains lead to improvements in the fraction *gauche* conformation even though the internal energy of the *gauche* vs. *trans* conformers based on butane was similar. This result emphasizes the relationship between the nonbond and internal portions of a force field and how proper optimization of both aspects are important for accurate treatment of atomic detail events is MD simulations. Further optimization of the hydrocarbon chains of the CHARMM27 lipid force field are ongoing (R. Pastor, D. Tobias and A.D. MacKerell, Jr., work in progress).

In addition to united and all-atom models are course-grained lipid models designed for simulations of extended lipid bilayers, unilamellar vesicles, and so on.[275,276] In such models the different moieties comprising the lipids are treated as extended spheres, as is water, with, in some cases, multiple waters being treated as a single sphere. For example, the head group in DPPC is modeled as one sphere each for the choline, phosphate, and glycerol moieties while each fatty acid chain includes a sphere for the ester linkage and four spheres for the aliphatic region. Interactions between particles involve LJ type interactions supplemented with electrostatic interactions between the head group moieties. These models allow for simulations of very large systems, corresponding to 1 million or more atoms, by significantly decreasing the actual number of particles, for extended time periods, such as microseconds, via the use of an integration time step of 50 fs. Although attractive for investigating large scale (both time and size!) phenomena in lipids, the majority of the atomic detail is lost. Further, it is necessary to carefully adjust the parameters for each system under study to ensure the behavior of the model in simulations is adequately representative of the experimental regimen.

Carbohydrates

Carbohydrates represent a unique challenge for empirical force fields for several reasons. With monosaccharides the conformational properties are dominated by a subtle balance of inter- and intramolecular hydrogen bonding involving the numerous hydroxyl groups and water. In addition, there are a significant number of different types of monosaccharides of biological interest. For example, the Pneumococcal 23-valent,[277] Meningococcal 4-valent,[278] *Salmonella typhi* Vi,[279] and Hib[280] vaccines contain a total of 24 monosaccharide/substituent combinations. These often contain functional groups that are important for molecular recognition, such as the acetamido, amino, sulfate, and uronic acid groups. At the polysaccharide level the complexity is further

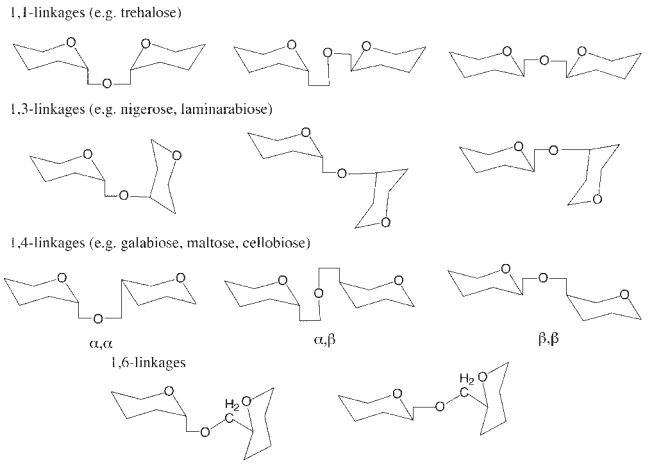

Figure 4. Example glycosyl linkages found in polysaccharides. Hydroxyl groups have been omitted for clarity.

amplified due to the presence of a variety of chemical connectivities in the form of glycosyl bonds compared to the repeating peptide bonds or phophodiester linkages in proteins and oligonucleotides, respectively. For example, there are 1,1; 1,2; 1,3; 1,4; 1,6; 2,2; and so on types of glycosyl linkages with each case having α and β anomers; a subset of the types of linkages in selected disaccharides is shown in Figure 4. Finally, the presence of multiple oxygens (or nitrogens) in carbohydrates allows for the delocalization of oxygen lone pairs into antibonding, σ^*, orbitals (i.e., hyperconjugation), which influences the conformational energies of the carbohydrates. These delocalizations are referred to as the *gauche*, anomeric, and exoanomeric effects.[281,282] Such complexity has lead to the development of a number of force fields for subsets of carbohydrates as described below; however, a comprehensive force field for this class of biomolecules is lacking.

A number of carbohydrate force fields have been developed for gas phase calculations. Recently, a MM4 force field for carbohydrates was reported.[283–286] This force field treats simple alcohols and ethers as well as a number of hexoses, although no tests on disaccharides have been preformed and, because the model was developed for the gas phase, its applicability for condensed phase simulations is unclear. Another example of a carbohydrate force field that is based on gas phase energies is that developed for the commercial CHARMm package[232] based on a variety of QM data. A variant of the commercial CHARMm force field, CHEAT, was designed to yield conformational properties that correspond to free energies in aqueous solution from gas phase calculations (i.e., a knowledge-based force field, see above).[287,288] In this model hydroxyl groups are treated as extended atoms, and the model was developed with the goal of obtaining reasonable estimates of polysaccharide conformations. Such gas phase-based force fields are useful for understanding general conformational properties of carbohydrates, although atomically detailed insights into the role of water, and their application in heterogeneous biomolecular simulations, may be expected to be minimal.

Several force fields associated with the program CHARMM are available that have been designed for use with an explicit solvent representation. Brady and coworkers have developed models for hexopyranose sugars based on small molecule geometric and vibrational data and on α-D-glucopyranose.[289] However, this force field yields incorrect conformational properties for the exocyclic hydroxyl,[290] glycosyl bonds are not treated, and the model is only limited to hydroxyl substituents. More recently, this model has been reoptimized, in part to be consistent with the CHARMM22 protein and CHARMM27 nucleic acid/lipid force fields and to yield improved conformational properties for the exocyclic hydroxyl group.[291,292] Other CHARMM-compatible force fields include one by Reiling et al.[293] that was based on a series of QM calculations on carbohydrate analogs and parameters for sulfates and sulfamates substitutents for use with β-D-glucose.[294]

A variety of carbohydrate simulations have employed the GROMOS force field,[221,295] which, as stated above, is a united atom model that is appropriate for simulations in aqueous solution. Variations include work by Ott and Meyer[296] that include corrections to more accurately treat the exo-anomeric effect. An all-atom revision has been proposed that accessed missing parameters from CHARMM and adjusted the LJ parameters based on condensed phase simulations.[288] Overall, these force fields yield reasonable results for monosaccharides, although studies on disaccharides have not been reported.

A significant number of carbohydrate force fields are available that are related to the AMBER force field. An early example extended AMBER to treat the acetamido group of GlcNAc as well as the glycosyl linkage.[297] A recent model has been developed based on vibrational fitting to glucose, although it is known to give the wrong anomer ratio in both solution and vacuum.[298] Glennon and Merz developed an hexapyronse force field in which semiempirical calculations were used to assign partial atomic charges to each atom (vs. all secondary hydroxyls having the same charge).[299,300] Woods and coworkers have developed the GLYCAM force field[301] again by considering the unique partial atomic charges for the individual atoms, including ensemble averaging in the charge determination[302,303] as well as additional optimization of internal parameters. Recent updates of the GLYCAM force field have included use of a 1,4 scale factor of 1.0[304] vs. 1/1.2 common to the remainder of the AMBER force fields and a general carbohydrate model that avoids anomeric carbon specific terms is available (R. Woods, personal communication). Such a change in the 1,4 factor may complicate studies of heterogeneous biomolecular systems. Another AMBER variant, AMB99C investigated 1–4 linkages in maltose and cyclodextrins by adjusting the appropriate parameters to reproduce QM conformational energies.[305,306] Kollman and coworkers produced a carbohydrate model that involved adjustment of torsional parameters for O–C–O–C and O–C–O–H connectivities based on small molecule QM data.[307] This model was shown to yield the correct ratio of the α vs. β anomers of glucose; however, the model currently is limited to monosaccharides and does not include moieties relevant to carbohydrates involved in molecular recognition. Other specialized AMBER variants include a united atom model that treats solvent via the Generalized Born/Solvent Accessible Area approach[308] and the SPACIBA force field that is optimized to produce accurate vibrational spectra via the use of Urey–Bradley–Shimanouchi spectroscopic terms that incorporate the effects of vibrational anharmonicity.[309] Again, these models appear to be rather specialized, such

that there applicability to a wide range of carbohydrates is questionable.

Carbohydrate models have also been developed based on extending the OPLS force field.[19] This work,[310] combined with recent refinements,[311] may represent the best force field available for simple carbohydrates. These efforts have been based on reoptimization of selected torsional parameters using QM conformational energies of D-glucopyranose, D-galactopyranose, D-mannopyranose, methyl D-glucopyranoside, and methyl D-mannopyranoside, along with emphasis on the properties of the exocyclic hydroxyl group.[310] However, these models deviate from standard OPLS by scaling the 1–5 and 1–6 electrostatic interactions to allow for treatment of vicinal hydroxyl groups. In addition, parameters for the various sugar derivatives and protein linkages for monosaccharides involved in molecular recognition are not available.

As is evident from the above discussion, a significant amount of effort has gone into the optimization of force fields for carbohydrates. However, no consensus force field for the wide range of monosaccharides, functional groups, and glycosyl linkages in polysaccharides is available. Given the essential role of carbohydrates in a wide variety of biological processes as well as problems associated with the experimental determination of carbohydrate structures,[312] this omission represents a significant gap in empirical force fields that is hindering our understanding the structure–activity relationships of carbohydrates.

Heterogeneous Biomolecular Systems

To date, the majority of empirical force field studies of biomolecules have involved "homogeneous" systems that included one type of biomolecule (e.g., protein or duplex DNA) along with, in some cases, solvent and ions. However, increased computational resources along with an ever increasing number of experimental structures for heterogeneous biomolecular systems that are central to biological processes require the application of empirical force fields to more complex systems. An important part of these applications will be the use of the proper force fields for the systems under study.

Biological phenomena at the atomic level are dominated by nonbonded interactions between molecules. Accordingly, when applying empirical force field-based methods to study heterogeneous biomolecular systems it is essential to have the nonbonded interactions between different aspects of the force field be properly balanced. To obtain such a balance it is generally necessary to use parameters that are part of the same force field; mixing of different force fields is ill advised, as the balance of nonbond interactions will be lost. Loss of such a balance can, for example, lead to an overestimation of protein–DNA interactions vs. protein–water and DNA–water interactions as well as possible leading to imbalances associated with DNA solvation being to favorable compared to protein solvation. Thus, when considering simulations of heterogeneous systems the range of molecules a force field covers must be taken into account. This consideration also applies to studies of small molecules interacting with biomolecules (e.g., drug–protein complexes, see below).

Currently, the most extensive biomolecular force fields are the CHARMM22 and CHARMM27 force fields. These include proteins, nucleic acids, lipids and, although limited, carbohydrates. To date, a number of successful studies have been performed on DNA–protein, DNA–lipid, and protein–lipid complexes using this model. The lack of availability of parameters for glycosyl linkages from carbohydrates to proteins or lipids currently hinders the application of CHARMM to glycoproteins and glycolipids. With AMBER studies on protein–nucleic acid and protein–carbohydrate systems would be appropriate. Care must be taken with the latter in cases where changes have been made in the treatment of nonbond parameters and 1,4 interactions (see above). It should be noted that the components of lipids are available in AMBER, although recent MD studies of lipids have not been reported. OPLS has been primarily used for simulations of proteins, although a significant amount of work with carbohydrates has been reported (see above), making this model appropriate for protein–carbohydrate systems. Care must be taken due to the recent introduction of 1,5 and 1,6 scaling factors in the carbohydrate model,[310,311] which may lead to compatibility problems with the proteins. With the GLYCAM and OPLS carbohydrate variants, the change of the form of the potential energy function via alterations of the scaling factors improves the ability of the models to treat carbohydrates, but may lead to problems when applied to heterogeneous systems. Also, the components for lipid simulations using OPLS are available, although no OPLS lipid simulations have been reported. The GROMOS united atom force fields is suited for protein–lipid and protein–carbohydrate simulations, while the lack of applications to nucleic acids makes its use in protein–nucleic acid studies less appropriate. Again, for the majority of these force fields parameters are often available for biomolecules to which the respective force fields are not typically applied. If an extension of any force fields is made to biomolecules not previously studied using that model, the simulator is advised to apply the respective force field to well-studied homogeneous model systems of the biomolecule being investigated to gauge the accuracy of the force field prior to applying it in more complex, heterogeneous systems.

Force Field Transferability: Application to Drug-Like Molecules

One of the major applications of empirical force fields is in the area of drug design and development.[313] For such applications it is necessary to have the relevant parameters for the drug-like molecules of interest. Given the huge dimensionality of chemical space, having the appropriate parameters for a wide range of compounds is not a trivial problem. Several force fields have addressed this problem directly. These include MMFF,[24,194] CVFF,[21,314] the commercial CHARMm force field,[232] CFF,[315] COMPASS,[316] the MM2/MM3/MM4 series,[16,283,317] UFF,[27] Dreiding,[26] the Tripos force field (Tripos, Inc.), among others. Typically, these force fields have been designed primarily to reproduce internal geometries, vibrations, and conformational energies, often sacrificing the quality of the nonbond interactions,[318] which are of obvious importance for accurately treating ligand–protein interactions. Exceptions are MMFF and COMPASS where nonbond parameters have

been investigated at a reasonable level of detail. In addition, the quality of these "wide-coverage" force fields in treating biomolecules is generally poorer than the more specialized biomolecular force fields. However, their wide coverage of chemical space often makes the use of these force fields desirable for screening of large databases of compounds. To account for limitations in these force fields in the treatment of, for example, proteins, it may be appropriate to restrain all of the protein except those residues in the vicinity of the ligand-binding site during minimization or MD studies of ligand–protein complexes.

Concerning the specialized biomolecular force fields, their transferability to drug-like molecules varies. For example, AMBER historically has been developed with transferability in mind, with recent efforts focused on more automated methods of parameter assignment.[319,320] CHARMM force fields have focused on more refined parameters, thereby limiting transferability. However, the modular approach used for the partial atomic charges (i.e., the sum of the atomic charges on moieties such as rings, carboxylates, etc., are integers) allows for the available parameters for the large collection of model compounds parameterized to date[45,321] to readily be combined into drug-like molecules, as previously described.[36] Work towards a more comprehensive, transferable force field that is compatible with the CHARMM biomolecular force fields is ongoing (B.R. Brooks and A.D. MacKerell, Jr., work in progress). A similar situation exists with OPLS where a wide variety of molecules that are often included as fragments in drug-like molecules have been explicitly parametrized.[40,322,323] Thus, with the biomolecular force fields more effort is typically required to set up and parameterize novel molecules; however, when performing calculations that require higher accuracy (e.g., free energy perturbation calculations for lead structure optimization), such efforts are typically warranted.

A study by Halgren nicely summarizes the extent of transferability of force fields.[194] In that study, a variety of force fields were tested for their ability to reproduce experimental data on the conformational energies for a variety of small compounds. While, based on that test set of compounds, some of the force fields worked well overall, in all cases there where catastrophic failures where the energetic ordering of conformers for a compound was incorrect. This emphasizes the need to carefully test a force field when it is being applied to molecules for which parameters were not explicitly developed. Such tests can be as simple a checking the relative energies of selected conformers against QM data (HF/6-31G* is accessible to most drug-like molecules and yields conformational energies that are representative for most molecules[194]) or be more detailed, involving careful checking of nonbonded interactions with the environment or condensed-phase simulations to determine experimentally accessible thermodynamic values. Again, the degree of accuracy required for a study dictates that amount of parameter testing and optimization required for novel molecules that are part of the study.

Summary

The quality of empirical force fields for biomolecules has increased to the level that the agreement with a variety of experimental data is often within the accuracy of the experimental method. This has allowed for empirical force field-based MD simulations on biomolecules to gain general acceptance in the scientific community. This acceptance, although by no means complete, is important as it allows for a more detailed picture of structure–function relationships in biomolecules, beyond those typically accessible to experiments, to be obtained. However, it is important for this confidence in force field-based methods to continue to develop, leading to wider acceptance of this important technique. This, in large part, is the responsibility of the individuals involved in force field development, who must continue to improve the accuracy of force fields in both the present form of the potential energy function as well as in extended forms, such as the 2D dihedral grid correction map method recently applied to proteins and, importantly, the inclusion of electronic polarizability in the potential energy function. Other advances will be based on algorithmic improvements that facilitate conformational sampling of biomolecules, thereby allowing for more rigorous comparisons with experiments.

Finally, the actual application of empirical force fields to a wide variety of biological systems will have a major impact on increasing the confidence in results from force field methods by the general scientific community. If the applications are performed in a carefully designed manner, using the appropriate methods and force fields, it can be anticipated that the quality of the agreement with the experiment will be enhanced, and that the atomic detail insights obtained from those simulations that are not currently accessible to experiment will stand the test of time. Such successes will ensure the growth of empirical force field methods in the biological sciences. It is hoped that the present review facilitates this growth.

Acknowledgments

Appreciation to my colleagues who have shared their thoughts on the accuracy, development and implementation of the CHARMM force fields and to Drs. Cheatham, Simmerling, Case, Wang, Tobias, Brady, Woods, and Feller for their helpful comments and suggestions on the present manuscript.

References

1. Karplus, M.; Petsko, G. A. Nature 1990, 347, 631.
2. Becker, O. M.; MacKerell, A. D., Jr.; Roux, B.; Watanabe, M., Eds.; Computational Biochemistry and Biophysics; Marcel-Dekker, Inc.: New York, 2001, p. 512.
3. Brust, J. C. Curr Neurol Neurosci Rep 2001, 1, 1.
4. Ewald, P. P. Ann Phys 1921, 64, 253.
5. Darden, T. In Computational Biochemistry and Biophysics; Becker, O. M., MacKerell, A. D., Jr., Roux, B., Watanabe, M., Eds.; Marcel Dekker, Inc.: New York, 2001, p. 91.
6. Beglov, D.; Roux, B. J Chem Phys 1994, 100, 9050.
7. Im, W.; Bernéche, S.; Roux, B. J Chem Phys 2001, 114, 2924.
8. Bishop, T. C.; Skeel, R. D.; Schulten, K. J Comp Chem 1997, 18, 1785.
9. Tuckerman, M.; Berne, B. J.; Martyna, G. J. J Chem Phys 1992, 97, 1990.

10. Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. J Chem Phys 1995, 103, 4613.
11. Martyna, G. J.; Tobias, D. J.; Klein, M. L. J Chem Phys 1994, 101, 4177.
12. Tuckerman, M. E.; Martyna, G. J. J Phys Chem B 2000, 104, 159.
13. Barth, E.; Schlick, T. J Chem Phys 1998, 109, 1633.
14. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. J Comput Chem 1983, 4, 187.
15. MacKerell, A. D., Jr.; Brooks, B.; Brooks, C. L., III; Nilsson, L.; Roux, B.; Won, Y.; Karplus, M. In Encyclopedia of Computational Chemistry; Schleyer, P. v. R.; Allinger, N. L.; Clark, T.; Gasteiger, J.; Kollman, P. A.; Schaefer, H. F., III; Schreiner, P. R., Eds.; John Wiley & Sons: Chichester, 1998, p. 271, vol. 1.
16. Burkert, U.; Allinger, N. L. Molecular Mechanics; American Chemical Society: Washington, DC, 1982, vol. 177.
17. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. J Am Chem Soc 1995, 117, 5179.
18. van Gunsteren, W. F. GROMOS. Groningen Molecular Simulation Program Package; University of Groningen: Groningen, 1987.
19. Jorgensen, W. L.; Tirado–Rives, J. J Am Chem Soc 1988, 110, 1657.
20. Lii, J.-L.; Allinger, N. L. J Comp Chem 1991, 12, 186.
21. Ewig, C. S.; Berry, R.; Dinur, U.; Hill, J.-R.; Hwang, M.-J.; Li, H.; Liang, C.; Maple, J.; Peng, Z.; Stockfisch, T. P.; Thacher, T. S.; Yan, L.; Ni, X.; Hagler, A. T. J Comp Chem 2001, 22, 1782.
22. Sun, H. J Phys Chem B 1998, 102, 7338.
23. Derreumaux, P.; Vergoten, G. J Chem Phys 1995, 102, 8586.
24. Halgren, T. A. J Comput Chem 1996, 17, 490.
25. Palmo, K.; Mannfors, B.; Mirkin, N. G.; Krimm, S. Biopolymers 2003, 68, 383.
26. Mayo, S. L.; Olafson, B. D.; Goddard, W. A., III. J Phys Chem 1990, 94, 8897.
27. Rappé, A. K.; Colwell, C. J.; Goddard, W. A., III; Skiff, W. M. J Am Chem Soc 1992, 114, 10024.
28. Blondel, A.; Karplus, M. J Comput Chem 1996, 17, 1132.
29. MacKerell, A. D., Jr.; Feig, M.; Brooks, C. L., III. J Am Chem Soc 2004, 126, 698.
30. MacKerell, A. D., Jr.; Feig, M.; Brooks, C. L., III. J Comp Chem 2004, 25, 1400.
31. Buckingham, A. D.; Fowler, P. W. Can J Chem 1985, 63, 2018.
32. Halgren, T. A. J Comput Chem 1996, 17, 520.
33. Halgren, T. A. J Am Chem Soc 1992, 114, 7827.
34. Reiher, W. E. Theoretical Studies of Hydrogen Bonding. Ph.D., Harvard University, 1985.
35. Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. J Am Chem Soc 1984, 106, 765.
36. MacKerell, A. D., Jr. In Computational Biochemistry and Biophysics; Becker, O. M., MacKerell, A. D., Jr., Roux, B., Watanabe, M., Eds.; Marcel Dekker, Inc.: New York, 2001, p. 7.
37. Jorgensen, W. L. J Phys Chem 1986, 90, 1276.
38. MacKerell, A. D., Jr.; Karplus, M. J Phys Chem 1991, 95, 10559.
39. Gough, C. A.; DeBolt, S. E.; Kollman, P. A. J Comp Chem 1992, 13, 963.
40. Jorgensen, W. L.; Maxwell, D. S.; Tirado–Rives, J. J Am Chem Soc 1996, 118, 11225.
41. Fox, T.; Kollman, P. A. J Phys Chem B 1998, 102, 8070.
42. Kaminski, G.; Duffy, E. M.; Matsui, T.; Jorgensen, W. L. J Phys Chem 1994, 98, 13077.
43. Yin, D.; MacKerell, A. D., Jr. J Comp Chem 1998, 19, 334.
44. Rizzo, R. C.; Jorgensen, W. L. J Am Chem Soc 1999, 121, 4827.
45. Chen, I.-J.; Yin, D.; MacKerell, A. D., Jr. J Comp Chem 2002, 23, 199.
46. Warshel, A.; Lifson, S. J Chem Phys 1970, 53, 582.
47. Hagler, A. T.; Maple, J. R.; Thacher, T. S.; Fitzgerald, G. B.; Dinur, U. In Computer Simulation of Biomolecular Systems; van Gunsteren, W. F., Weiner, P. K., Eds.; ESCOM: Leiden, 1989, p. 149.
48. Foloppe, N.; MacKerell, A. D., Jr. J Comp Chem 2000, 21, 86.
49. Shirts, M. R.; Pitera, J. W.; Swope, W. C.; Pande, V. S. J Chem Phys 2003, 119, 5740.
50. Lamoureux, G.; MacKerell, A. D., Jr.; Roux, B. J Chem Phys 2003, 119, 5185.
51. Halgren, T. A.; Damm, W. Curr Opin Struct Biol 2001, 11, 236.
52. Ponder, J. W.; Case, D. A. Adv Protein Chem 2003, 66, 27.
53. Rick, S. W.; Stuart, S. J. Rev Comp Chem 2002, 18, 89.
54. Caldwell, J.; Dang, L. X.; Kollman, P. A. J Am Chem Soc 1990, 112, 9144.
55. Sprik, M.; Klein, M. L. J Chem Phys 1988, 89, 7556.
56. Wallqvist, A.; Berne, B. J. J Phys Chem 1993, 97, 13841.
57. Bernardo, D. N.; Ding, Y.; Krogh–Jespersen, K.; Levy, R. M. J Phys Chem 1994, 98, 4180.
58. Dang, L. X. J Phys Chem B 1998, 102, 620.
59. Rick, S. W.; Stuart, S. J.; Bader, J. S.; Berne, B. J. J Mol Liq 1995, 66/66, 31.
60. Rick, S. W.; Berne, B. J. J Am Chem Soc 1996, 118, 672.
61. Bryce, R. A.; Vincent, M. A.; Malcolm, N. O. J.; Hillier, I. H.; Burton, N. A. J Chem Phys 1998, 109, 3077.
62. Yoshii, N.; Miyauchi, R.; Niura, S.; Okazaki, S. Chem Phys Lett 2000, 317, 414.
63. Asensio, J. L.; Canada, F. J.; Cheng, X.; Khan, N.; Mootoo, D. R.; Jimenez–Barbero, J. Chemistry 2000, 6, 1035.
64. Llanta, E.; Ando, K.; Rey, R. J Phys Chem B 2001, 105, 7783.
65. Patel, S.; Brooks, C. L., III. J Comput Chem 2004, 25, 1.
66. Stern, H. A.; Kaminski, G. A.; Banks, J. L.; Zhou, R.; Berne, B. J.; Friesner, R. A. J Phys Chem B 1999, 103, 4730.
67. Mannfors, B.; Palmo, K.; Krimm, S. J Mol Struct 2000, 556, 1.
68. Dick, B. G., Jr.; Overhauser, A. W. Phys Rev 1958, 112, 90.
69. Pratt, L. R. Mol Phys 1980, 40, 347.
70. Stuart, S. J.; Berne, B. J. J Phys Chem 1996, 100, 11934.
71. van Marren, P. J.; van der Spoel, D. J Phys Chem B 2001, 105, 2618.
72. Lamoureux, G.; Roux, B. J Chem Phys 2003, 119, 3025.
73. Chen, B.; Xing, J.; Siepmann, I. J. J Phys Chem B 2000, 104, 2391.
74. Stern, H. A.; Rittner, F.; Berne, B. J.; Friesner, R. A. J Chem Phys 2001, 115, 2237.
75. Ren, P.; Ponder, J. W. J Phys Chem B 2003, 107, 5933.
76. Grossfield, A.; Ren, P.; Ponder, J. W. J Am Chem Soc 2003, 125, 15671.
77. Shelley, J. C.; Sprik, M.; Klein, M. L. Langmuir 1993, 9, 916.
78. Gao, J.; Habibollazadeh, D.; Shao, L. J Phys Chem 1995, 99, 16460.
79. Caldwell, J. W.; Kollman, P. A. J Phys Chem 1995, 99, 6208.
80. Freindorf, M.; Gao, J. J Comp Chem 1996, 17, 386.
81. Cieplak, P.; Caldwell, J. W.; Kollman, P. A. J Comp Chem 2001, 22, 1048.
82. Caldwell, J. W.; Kollman, P. A. J Am Chem Soc 1995, 117, 4177.
83. Dang, L. X. J Phys Chem B 1999, 103, 8195.
84. Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A.; Cao, Y. X.; Murphy, R. B.; Zhou, R.; Halgren, T. A. J Comp Chem 2002, 23, 1515.
85. Patel, S.; MacKerell, A. D., Jr.; Brooks, C. L., III. J Comp Chem 2004, 25, 1504.
86. Anisimov, V. M.; Vorobyov, I. V.; Lamoureux, G.; Noskov, S.; Roux, B.; MacKerell, A. D., Jr. Biophys J 2004, 86, 415a.
87. Morita, A.; Kato, S. J Chem Phys 1999, 110, 11987.
88. Jorgensen, W. L.; Tirado–Rives, J. J Am Chem Soc 1988, 110, 1657.
89. Stillinger, F. H.; Rahman, A. J Chem Phys 1974, 60, 1545.
90. Dixon, R. W.; Kollman, P. A. J Comp Chem 1997, 18, 1632.

91. Mahoney, M. W.; Jorgensen, W. L. J Chem Phys 2000, 112, 8910.
92. Mahoney, M. W.; Jorgensen, W. L. J Chem Phys 2001, 114, 363.
93. Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. J Phys Chem 1993, 97, 10269.
94. Burley, S. K.; Petsko, G. A. Science 1985, 229, 23.
95. Burley, S. K.; Petsko, G. A. Adv Protein Chem 1988, 39, 125.
96. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. J Chem Phys 1983, 79, 926.
97. Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. J Phys Chem 1987, 91, 6269.
98. Levitt, M.; Hirshberg, M.; Sharon, R.; Laidig, K. E.; Daggett, V. J Phys Chem B 1997, 101, 5051.
99. Feller, S. E.; Pastor, R. W.; Rojnuckarin, A.; Bogusz, S.; Brooks, B. R. J Phys Chem 1996, 100, 17011.
100. Glättli, A.; Daura, X.; van Gunsteren, W. F. J Comp Chem 2003, 24, 1087.
101. Davis, M. E.; McCammon, J. A. Chem Rev 1990, 90, 509.
102. Honig, B. J Phys Chem 1993, 97, 1101.
103. Feig, M.; Brooks, C. L., III. Curr Opin Struct Biol 2004, 14, 217.
104. Wesson, L.; Eisenberg, D. Protein Sci 1992, 1, 227.
105. Gilson, M. K.; Honig, B. Proteins 1988, 4, 7.
106. Nina, M.; Beglov, D.; Roux, B. J Phys Chem B 1997, 101, 5239.
107. Banavali, N. K.; Roux, B. J Phys Chem B 2002, 106, 11026.
108. Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. J Am Chem Soc 1990, 112, 6127.
109. Schaefer, M.; Karplus, M. J Phys Chem 1996, 100, 1578.
110. Jayaram, B.; Sprous, D.; Beveridge, D. L. J Phys Chem B 1998, 102, 9571.
111. Onufriev, A.; Bashford, D.; Case, D. A. J Phys Chem B 2000, 104, 3712.
112. Zhang, L. Y.; Gallicchio, E.; Friesner, R. A.; Levy, R. M. J Comp Chem 2001, 22, 591.
113. Lee, M. S.; Feig, M.; Salsbury, F. R., Jr.; Brooks, C. L., III. J Comp Chem 2003, 24, 1348.
114. Qui, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. J Phys Chem A 1997, 101, 3005.
115. Gallicchio, E.; Zhang, L. Y.; Levy, R. M. J Comp Chem 2003, 23, 517.
116. Gallicchio, E.; Levy, R. M. J Comp Chem 2004, 25, 479.
117. Florián, J.; Warshel, A. J Phys Chem B 1997, 101, 5583.
118. Lazaridis, T.; Karplus, M. Proteins 1999, 35, 133.
119. Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E., III. Acc Chem Res 2000, 33, 889.
120. Jayaram, B.; McConnell, K. J.; Dixit, S. B.; Beveridge, D. L. J Comp Chem 2002, 23, 1.
121. Gohlke, H.; Kiel, C.; Case, D. J Mol Biol 2003, 330, 891.
122. Ferrara, P.; Gohlke, H.; Price, D. J.; Klebe, G.; Brooks, C. L., III. J Med Chem 2004, 47, 3032.
123. Steinbach, P. J.; Brooks, B. R. J Comput Chem 1994, 15, 667.
124. Allen, M. P.; Tildesley, D. J. Computer Simulation of Liquids; Clarendon Press: Oxford, 1987.
125. Lague, P.; Pastor, R. W.; Brooks, B. R. J Phys Chem B 2004, 108, 363.
126. Stote, R. H.; States, D. J.; Karplus, M. J Chim Phys 1991, 88, 2419.
127. Ding, H.-Q.; Karasawa, N.; Goddard, W. A., III. J Chem Phys 1992, 97, 4309.
128. White, C. A.; Head–Gordon, M. J Chem Phys 1994, 101, 6593.
129. Darden, T.; York, D.; Pedersen, L. J Chem Phys 1993, 98, 10089.
130. Kuwajima, S.; Warshel, A. J Chem Phys 1988, 89, 3751.
131. Toukmaji, A.; Board, J. A. Comp Phys Commun 1996, 95, 78.
132. Pollock, E.; Glosli, J. Comp Phys Commun 1996, 95, 93.
133. Figueirido, F.; Del Buono, G. S.; Levy, R. M. J Chem Phys 1995, 103, 6133.
134. de Souza, O. N.; Ornstein, R. L. Biophys J 1997, 72, 2395.
135. Cheatham, T. E., III; Young, M. A. Biopolymers 2001, 56, 232.
136. Berkowitz, M. L.; McCammon, J. A. Chem Phys Lett 1982, 90, 215.
137. Brooks, C. L., III; Karplus, M. J Chem Phys 1983, 79, 6312.
138. Berneche, S.; Roux, B. Nature 2001, 414, 73.
139. Huang, N.; Banavali, N. K.; MacKerell, A. D., Jr. Proc Natl Acad Sci USA 2003, 100, 68.
140. Pulay, P.; Fogarasi, G.; Pang, F.; Boggs, J. E. J Am Chem Soc 1979, 101, 2550.
141. Brooks, C. L., III; Case, D. A. Chem Rev 1993, 93, 2487.
142. Beachy, M. D.; Chasman, D.; Murphy, R. B.; Halgren, T. A.; Friesner, R. A. J Am Chem Soc 1997, 119, 5908.
143. Wieczorek, R.; Dannenberg, J. J. J Am Chem Soc 2003, 125, 14065.
144. Foloppe, N.; Nilsson, L.; MacKerell, A. D., Jr. Biopolymers (Nucleic Acid Sci) 2002, 61, 61.
145. Shishkin, O. V.; Gorb, L.; Zhikol, O. A.; Leszczynski, J. J Biomol Struct Dyn 2004, 21, 537.
146. MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E., III.; Roux, B.; Schlenkrich, M.; Smith, J.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz–Kuczera, J.; Yin, D.; Karplus, M. J Phys Chem B 1998, 102, 3586.
147. Foloppe, N.; MacKerell, A. D., Jr. J Phys Chem B 1998, 102, 6669.
148. Foloppe, N.; MacKerell, A. D., Jr. Biophys J 1999, 76, 3206.
149. Foloppe, N.; MacKerell, A. D., Jr. J Phys Chem B 1999, 103, 10955.
150. Berman, H. M.; Olson, W. K.; Beveridge, D. L.; Westbrook, J.; Gelbin, A.; Demeny, T.; Hsieh, S.-H.; Srinivasan, A. R.; Schneider, B. Biophys J 1992, 63, 751.
151. Bush, B. L.; Bayly, C. I.; Halgren, T. A. J Comp Chem 1999, 20, 1495.
152. Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. J Comp Chem 2000, 21, 132.
153. Gilson, M. K.; Gilson, H. S.; Potter, M. J. J Chem Inf Comp Sci 2003, 43, 1982.
154. Singh, U. C.; Kollman, P. A. J Comp Chem 1984, 5, 129.
155. Chirlian, L. E.; Francl, M. M. J Comput Chem 1987, 8, 894.
156. Henchman, R. H.; Essex, J. W. J Comp Chem 1999, 20, 483.
157. Merz, K. M. J Comput Chem 1992, 13, 749.
158. Francl, M. M.; Carey, C.; Chirlian, L. E.; Gange, D. M. J Comp Chem 1996, 17, 367.
159. Laio, A.; VandeVondele, J.; Rothlisberger, U. J Phys Chem B 2002, 106, 7300.
160. Cieplak, P.; Cornell, W. D.; Bayly, C. I.; Kollman, P. K. J Comp Chem 1995, 16, 1357.
161. MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack Jr., R. L.; Evanseck, J.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, I., W. E.; Roux, B.; Schlenkrich, M.; Smith, J.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz–Kuczera, J.; Yin, D.; Karplus, M. J Phys Chem B 1998, 102, 3586.
162. Jorgensen, W. L. J Am Chem Soc 1984, 106, 6638.
163. MacKerell, A. D., Jr.; Wiórkiewicz–Kuczera, J.; Karplus, M. J Am Chem Soc 1995, 117, 11946.
164. Yin, D.; MacKerell, A. D., Jr. J Phys Chem 1996, 100, 2588.
165. Schlenkrich, M.; Brickmann, J.; MacKerell, A. D., Jr.; Karplus, M. In Biological Membranes: A Molecular Perspective from Computation and Experiment; Merz, K. M., Roux, B., Eds.; Birkhaüser: Boston, 1996, p. 31.

166. McCammon, J. A.; Gelin, B. R.; Karplus, M. Nature 1977, 267, 585.
167. Kaminski, G.; Friesner, R. A.; Tirado–Rives, J.; Jorgensen, W. L. J Phys Chem B 2001, 105, 6474.
168. Price, D. J.; Brooks, C. L., III. J Comp Chem 2002, 23, 1045.
169. Ramachandran, G. N.; Ramakrishnan, C.; Sasisekharan, V. J Mol Biol 1963, 7, 95.
170. Head–Gordon, T.; Head–Gordon, M.; Frisch, M. J.; Brooks, C. L.; Pople, J. A. J Am Chem Soc 1991, 113, 5989.
171. Ono, S.; Kuroda, M.; Higo, J.; Nakajima, N.; Nakamura, H. J Comp Chem 2002, 23, 470.
172. Vargas, R.; Garza, J.; Hay, B. P.; Dixon, D. A. J Phys Chem A 2002, 106, 3213.
173. Dunbrack, R. L., Jr.; Cohen, F. E. Protein Sci 1997, 6, 1661.
174. Dunbrack, R. L., Jr. Culledpdb: Non-redundant set of protein sidechains from the PDB; http://www.fccc.edu/research/labs/dunbrack/culledpdb.html ed. Philadelphia, 2002.
175. Liu, H.; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Hermans, J.; Yang, W. Proteins 2001, 44, 484.
176. Hu, H.; Elstner, M.; Hermans, J. Proteins Struct Funct Genet 2003, 50, 451.
177. Feig, M.; MacKerell, A. D., Jr.; Brooks, C. L., III. J Phys Chem B 2002, 107, 2831.
178. Morgan, D. M.; Lynn, D. G.; Miller–Aure, H.; Meredith, S. C. Biochemistry 2001, 40, 14020.
179. Villa, A.; Mark, A. E. J Comp Chem 2002, 23, 548.
180. MacCallum, J. L.; Tieleman, P. J Comp Chem 2003, 24, 1930.
181. Deng, Y.; Roux, B. Submitted.
182. Garcia, A. E.; Sanbonmatsu, K. Y. Proc Natl Acad Sci USA 2002, 99, 2782.
183. Okur, A.; Strockbine, B.; Hornak, V.; Simmerling, C. J Comput Chem 2003, 24, 21.
184. Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Ceiplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. J Comp Chem 2003, 24, 1999.
185. Tsui, V.; Case, D. A. Biopolymers 2000, 56, 275.
186. van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hünenberger, P. H.; Krüger, P.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G. Biomolecular Simulation: The GROMOS96 Manual and User Guide; BIOMOS b.v.: Zürich, 1996.
187. Neria, E.; Fischer, S.; Karplus, M. J Chem Phys 1996, 105, 1902.
188. Daura, X.; Hünenberger, P. H.; Mark, A. E.; Querol, E.; Avilés, F. X.; van Gunsteren, W. F. J Am Chem Soc 1996, 118, 6285.
189. Schuler, L. D.; Daura, X.; van Gunsteren, W. F. J Comp Chem 2001, 22, 1205.
190. Schaefer, M.; Bartels, C.; LeClerc, F.; Karplus, M. J Comp Chem 2001, 22, 1857.
191. Lee, M. S.; Salsbury, F. R., Jr.; Brooks, C. L., III. J Chem Phys 2002, 116, 10606.
192. Im, W.; Lee, M. S.; Brooks, C. L., III. J Comput Chem 2003, 24, 1691.
193. Ferrara, P.; Apostolakis, J.; Caflisch, A. Proteins 2002, 46, 24.
194. Halgren, T. A. J Comp Chem 1999, 20, 730.
195. Ferro, D. R.; McQueen, J. E.; McCown, J. T.; Hermans, J. J Mol Biol 1980, 136, 1.
196. Hermans, J.; Berendsen, H. J. C.; van Gunsteren, W. F.; Postma, J. P. M. Biopolymers 1984, 23, 1513.
197. Levitt, M. ENCAD—Energy Calculations and Dynamics; Molecular Applications Group: Stanford, CA, and Rehovot, Israel, 1990.
198. Levitt, M.; Hirshberg, M.; Sharon, R.; Daggett, V. Comp Phys Commun 1995, 91, 215.
199. McQuarrie, D. A. Statistical Mechanics; Harper & Row: New York, 1976.
200. Momany, F. A.; Carruthers, L. M.; McGuire, R. F.; Scheraga, H. A. J Phys Chem 1974, 78, 1595.
201. Momany, F. A.; McGuire, R. F.; Burgess, A. W.; Scheraga, H. A. J Phys Chem 1975, 79, 2361.
202. Némethy, G.; Pottle, M. S.; Scheraga, H. A. J Phys Chem 1983, 87, 1883.
203. Némethy, G.; Gibson, K. D.; Palmer, K. A.; Yoon, C. N.; Paterlini, G.; Zagari, A.; Rumsey, S.; Scheraga, H. A. J Phys Chem 1992, 96, 6472.
204. Liwo, A.; Oldziej, S.; Pincus, M. R.; Wawak, R. J.; Rackovsky, S.; Scheraga, H. A. J Comp Chem 1997, 18, 849.
205. Lee, J.; Park, K.; Lee, J. J Phys Chem B 2002, 106, 11647.
206. Maksimiak, K.; Rodziewicz–Motowidlo, S.; Czaplewski, C.; Liwo, A.; Scheraga, H. A. J Phys Chem B 2003, 107, 13496.
207. Mirny, L. A.; Shakhnovich, E. I. J Mol Biol 1996, 264, 1164.
208. Lemak, A. S.; Gunn, J. R. J Phys Chem B 2000, 104, 1097.
209. Hassinen, T.; Peräkylä, M. J Comp Chem 2001, 22, 1229.
210. Lu, H.; Skolnick, J. Proteins 2001, 44, 223.
211. Bentacourt, M. R. Proteins 2003, 53, 889.
212. Russ, W. P.; Ranganathan, R. Curr Opin Struct Biol 2002, 12, 447.
213. Venclovas, C.; Zemla, A.; Fidelis, K.; Moult, J. Proteins 2001, Suppl 5, 163.
214. MacKerell, A. D., Jr.; Nilsson, L. In Computational Biochemistry and Biophysics; Becker, O. M., MacKerell, A. D., Jr., Roux, B., Watanabe, M., Eds.; Marcel Dekker, Inc.: New York, 2001, p. 441.
215. Levitt, M. Cold Spring Harbor Symp Quant Biol 1983, 47, 271.
216. Tidor, B.; Irikura, K. K.; Brooks, B. R.; Karplus, M. J Biomol Struct Dyn 1983, 1, 231.
217. Singh, U. C. Proc Natl Acad Sci USA 1985, 82, 755.
218. Srinivasan, J.; Withka, J. M.; Beveridge, D. L. Biophys J 1990, 58, 533.
219. Weiner, P. K.; Kollman, P. A. J Comput Chem 1981, 2, 287.
220. Nilsson, L.; Karplus, M. J Comput Chem 1986, 7, 591.
221. van Gunsteren, W. F.; Berendsen, H. J. C. GROMOS 86: Groningen Molecular Simulation Program Package; University of Groningen: Groningen, The Netherlands, 1986.
222. York, D. M.; Yang, W.; Lee, H.; Darden, T.; Pedersen, L. G. J Am Chem Soc 1995, 117, 5001.
223. Cheatham, T. E., III; Miller, J. L.; Fox, T.; Darden, T. A.; Kollman, P. A. J Am Chem Soc 1995, 117, 4193.
224. Weerasinghe, S.; Smith, P. E.; Mohan, V.; Cheng, Y.-K.; Pettitt, B. M. J Am Chem Soc 1995, 117, 2147.
225. Norberg, J.; Nilsson, L. J Chem Phys 1996, 104, 6052.
226. MacKerell, A. D., Jr. J Phys Chem B 1997, 101, 646.
227. Norberg, J.; Nilsson, L. Biophys J 2000, 79, 1537.
228. MacKerell, A. D., Jr.; Banavali, N. B.; Foloppe, N. Biopolymers 2001, 56, 257.
229. MacKerell, A. D., Jr.; Banavali, N. K. J Comp Chem 2000, 21, 105.
230. Cheatham, T. E., III; Cieplak, P.; Kollman, P. A. J Biomol Struct Dyn 1999, 16, 845.
231. Langley, D. R. J Biomol Struct Dyn 1998, 16, 487.
232. Momany, F. A.; Rone, R. J Comput Chem 1992, 13, 888.
233. Manning, G. S. Q Rev Biophys 1978, 11, 179.
234. Record, M. T., Jr.; Anderson, C. F.; Lohman, T. M. Q Rev Biophys 1978, 11, 103.
235. Beveridge, D. L.; McConnell, K. J. Curr Opin Struct Biol 2000, 10, 182.
236. Jóhannesson, H.; Halle, B. J Am Chem Soc 1998, 120, 6859.
237. Chiu, T. K.; Kaczor–Grzeskowiak, M.; Dickerson, R. E. J Mol Biol 1999, 292, 589.
238. McFail–Isom, L.; Sines, C. C.; Williams, L. D. Curr Opin Struct Biol 1999, 9, 298.
239. McConnell, K. J.; Beveridge, D. L. J Mol Biol 2000, 304, 803.

240. Hamelberg, D.; Williams, L. D.; Wilson, W. D. J Am Chem Soc 2001, 123, 7745.
241. Calladine, C. R.; Drew, H. R. Understanding DNA: The Molecule and How it Works; Academic Press: New York, 1997.
242. Lebrun, A.; Lavery, R. Curr Opin Struct Biol 1997, 7, 348.
243. Banavali, N. K.; MacKerell, A. D., Jr. J Mol Biol 2002, 319, 141.
244. Feig, M.; Zacharias, M.; Pettitt, B. M. Biophys J 2001, 81, 353.
245. Pan, Y.; MacKerell, A. D., Jr. Nucleic Acid Res 2003, 31, 7131.
246. Norberg, J.; Nilsson, L. J Phys Chem 1996, 100, 2550.
247. Kéblová, K.; Spackova, N.; Sponer, J. E.; Koca, J.; Sponer, J. Nucleic Acid Res 2003, 31, 6942.
248. Sarzynska, J.; Kulinski, T.; Nilsson, L. Biophys J 2000, 79, 1213.
249. Sarzynska, J.; Nilsson, L.; Kulinski, T. Biophy J 2003, 85, 3445.
250. Brameld, K.; Dasgupta, S.; Goddard, W. A., III. J Phys Chem B 1997, 101, 4851.
251. Ryjacek, F.; Kubar, T.; Hobza, P. J Comp Chem 2003, 24, 1891.
252. Evanseck, J. D.; Karplus, M. Unpublished work. 1994.
253. Tapia, O.; Velázquez, I. J Am Chem Soc 1997, 119, 5934.
254. Pranata, J.; Wierschke, S. G.; Jorgensen, W. L. J Am Chem Soc 1991, 113, 2810.
255. Flatters, D.; Zakrzewska, K.; Lavery, R. J Comp Chem 1997, 18, 1043.
256. Harvey, S. C.; Wang, C.; Teletchea, S.; Lavery, R. J Comput Chem 2003, 24.
257. Sundaralingam, M. Ann NY Acad Sci 1972, 195, 324.
258. Saiz, L.; Klein, M. L. Acc Chem Res 2002, 35, 482.
259. Pastor, R. W.; Venable, R. M.; Feller, S. E. Acc Chem Res 2002, 35, 438.
260. Daura, X.; Mark, A. E.; van Gunsteren, W. F. J Comp Chem 1998, 19, 535.
261. Chiu, S. W.; Clark, M. M.; Jakobsson, E.; Subramaniam, S.; Scott, H. L. J Phys Chem B 1999, 103, 6323.
262. Chiu, S. W.; Jakobsson, E.; Subramaniam, S.; Scott, H. L. Biophy J 1999, 77, 2462.
263. Ryckaert, J. P.; Bellemans, A. Chem Phys Lett 1975, 30, 123.
264. Berger, O.; Edholm, O.; Jahnig, F. Biophys J 1997, 72, 2002.
265. Anézo, C.; de Vries, A. H.; Höltje, H.-D.; Tieleman, P.; Marrink, S.-J. J Phys Chem B 2003, 107, 9424.
266. Patra, M.; Karttunen, M.; Hyvönen, M. T.; Falck, E.; Lindqvist, P.; Vattulainen, I. Biophys J 2003, 84, 3636.
267. Smondyrev, A. M.; Berkowitz, M. L. Biophys J 2000, 78, 1672.
268. Feller, S. E.; Yin, D.; Pastor, R. W.; MacKerell, A. D., Jr. Biophys J 1997, 73, 2269.
269. Feller, S. E.; Gawrisch, K.; MacKerell, A. D., Jr. J Am Chem Soc 2002, 124, 318.
270. Venable, R. M.; Zhang, Y.; Hardy, B. J.; Pastor, R. W. Science 1993, 262, 223.
271. MacKerell, A. D., Jr. J Phys Chem 1995, 99, 1846.
272. Woolf, T. B.; Roux, B. Proc Natl Acad Sciences USA 1994, 91, 11631.
273. Tobias, D. J.; Tu, K.; Klein, M. L. Curr Opin Colloid Interface Sci 1997, 2, 15.
274. Smith, G. D.; Jaffe, R. L. J Phys Chem 1996, 100, 18718.
275. Shelley, J. C.; Shelley, M. Y.; Reeder, R. C.; Bandyopadhyay, S.; Moore, P. B.; Klein, M. L. J Phys Chem B 2001, 105, 9785.
276. Marrink, S.-J.; Mark, A. E. J Am Chem Soc 2003, 125, 15233.
277. Abeygunawardana, C.; Williams, T. C.; Summer, J. S.; Hennessey, J. P. Anal Biochem 2000, 297, 266.
278. Frasch, C. In Immunization Against Bacterial Disease; Easmon, C. S. F., Jeljaszewicz, J., Eds.; Academic Press: London, 1983, p. 115, vol. 2.
279. Lemercinier, X.; Martinez–Cabrera, I.; Jones, C. Biologicals 2000, 28, 17.
280. Egan, W.; Schneerson, R.; Werner, K. E.; Zon, G. J Am Chem Soc 1982, 104, 2898.
281. Wolfe, S. Acc Chem Res 1972, 5, 102.
282. Wolfe, S.; Whangbo, M.-H.; Mitchell, D. J. Carbohydr Res 1979, 69, 1.
283. Allinger, N. L.; Chen, K. H.; Lii, J. H.; Durkin, K. A. J Comput Chem 2003, 24, 1447.
284. Lii, J. H.; Chen, K. H.; Allinger, N. L. J Comput Chem 2003, 24, 1504.
285. Lii, J. H.; Chen, K. H.; Durkin, K. A.; Allinger, N. L. J Comput Chem 2003, 24, 1473.
286. Lii, J. H.; Chen, K. H.; Grindley, T. B.; Allinger, N. L. J Comput Chem 2003, 24, 1490.
287. Grootenhuis, P. D. J.; Haasnoot, C. A. G. Mol Simulat 1993, 10, 75.
288. Kouwijzer, M. L. C. E.; van Eijck, B. P.; Kooijman, H.; Kroon, J. Acta Crystallogr 1995, B51, 209.
289. Ha, S. N.; Giammona, A.; Field, M.; Brady, J. W. Carbohydr Res 1988, 180, 207.
290. Kouwijzer, M. L. C. E.; van Eijck, B. P.; Kroes, S. J.; Kroon, J. J Comp Chem 1993, 14, 1281.
291. Palma, R.; Zuccato, P.; Himmel, M. E.; Liang, G.; Brady, J. W. In Glycosyl Hydrolases in Biomass Conversion; Himmel, M. E., Ed.; American Chemical Society: Washington, DC, 2000, p. 112.
292. Kuttel, M.; Brady, J. W.; Naidoo, K. J. J Comp Chem 2002, 23, 1236.
293. Reiling, S.; Schlenkrich, M.; Brickmann, J. J Comp Chem 1996, 17, 450.
294. Huige, C. J. M.; Altona, C. J Comput Chem 1995, 16, 56.
295. Scott, W. R. P.; Hünenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Krüger, P.; van Gunsteren, W. F. J Phys Chem A 1999, 103, 3596.
296. Ott, K.-H.; Meyer, B. J Comput Chem 1996, 17, 1068.
297. Homans, S. W. Biochemistry 1990, 29, 9110.
298. Gregurick, S. K.; Liu, J. H.-Y.; Brant, D. A.; Gerber, R. B. J Phys Chem B 1999, 103, 3476.
299. Glennon, T. M.; Merz, K. M. J Mol Struct (Theochem) 1997, 395, 157.
300. Glennon, T. M.; Zheng, Y.-J.; Grand, S. M. L.; Shutzberg, B. A.; Merz, K. M. J Comput Chem 1994, 15, 1019.
301. Woods, R. J.; Dwek, R. A.; Edge, C. J.; Fraser–Reid, B. J Phys Chem 1995, 99, 3832.
302. Woods, R. J.; Chappelle, R. J Mol Struct (Theochem) 2000, 527, 149.
303. Basma, M.; Sundara, S.; Calgan, D.; Vernali, T.; Woods, R. J. J Comput Chem 2001, 22, 1125.
304. Kirschner, K. N.; Woods, R. J. Proc Natl Acad Sci USA 2001, 98, 10541.
305. Momany, F. A.; Willett, J. L. Carbohydr Res 2000, 326, 194.
306. Momany, F. A.; Willett, J. L. Carbohydr Res 2000, 326, 210.
307. Simmerling, C.; Fox, T.; Kollman, P. A. J Am Chem Soc 1998, 120, 5771.
308. Senderowitz, H.; Parish, C.; Still, W. C. J Am Chem Soc 1996, 118, 2078.
309. Durier, V.; Tristram, F.; Vergoten, G. J Mol Struct (Theochem) 1997, 395–396, 81.
310. Damm, W.; Frontera, A.; Tirado–Rives, J.; Jorgensen, W. L. J Comput Chem 1997, 18, 1955.
311. Kony, D.; Damm, W.; Stoll, S.; van Gunsteren, W. F. J Comput Chem 2002, 23, 1416.
312. Rao, V. S. R.; Qasbo, P. K.; Balaji, P. V.; Chandrasekaran, R. Conformation of Carbohydrates; Harwood Academic Publishers: Canada, 1998.
313. Wong, C. F.; McCammon, J. A. Adv Protein Chem 2003, 66, 87.
314. Lifson, S.; Hagler, A. T.; Dauber, P. J Am Chem Soc 1979, 101, 5111.

315. Hwang, M. J.; Stockfisch, T. P.; Hagler, A. T. J Am Chem Soc 1994, 116, 2515.
316. Sun, H. J Phys Chem B 1998, 102, 7338.
317. Allinger, N. L.; Yuh, Y. H.; Lii, J.-L. J Am Chem Soc 1989, 111, 8551.
318. Kaminski, G.; Jorgensen, W. L. J Phys Chem 1996, 100, 18010.
319. Wang, J.; Cieplak, P.; Kollman, P. A. J Comp Chem 2000, 21, 1049.
320. Wang, J.; Kollman, P. A. J Comp Chem 2001, 22, 1219.
321. Yin, D. Parametrization for Empirical Force Field Calculations & A Theoretical Study of Membrane Permeability of Pyridine Derivatives. Ph.D., University of Maryland, 1997.
322. McDonald, N. A.; Jorgensen, W. L. J Phys Chem B 1998, 102, 8049.
323. Price, M. L. P.; Ostrovsky, D.; Jorgensen, W. L. J Comp Chem 2001, 22, 1340.
324. Guo, H.; Karplus, M. J Phys Chem 1992, 96, 7273.
325. Kitano, M.; Fukuyama, T.; Kuchitsu, B. Bull Chem Soc Jpn 1973, 46, 384.
326. Hamzaoui, F.; Baert, F. Acta Crystallogr C 1994, C50, 757.