Biochimica et Biophysica Acta 1850 (2015) 861–871

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbagen

Review

CHARMM additive and polarizable force fields for biophysics and computer-aided drug design☆

K. Vanommeslaeghe, A.D. MacKerell Jr. *

Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, Baltimore, MD 21201, USA

ARTICLE INFO

Article history:
Received 7 July 2014
Received in revised form 10 August 2014
Accepted 12 August 2014
Available online 19 August 2014

Keywords:
Molecular dynamics
Empirical force field
Potential energy function
Molecular Mechanics
Computer-aided drug design
Biophysics

ABSTRACT

Background: Molecular Mechanics (MM) is the method of choice for computational studies of biomolecular systems owing to its modest computational cost, which makes it possible to routinely perform molecular dynamics (MD) simulations on chemical systems of biophysical and biomedical relevance.
Scope of review: As one of the main factors limiting the accuracy of MD results is the empirical force field used, the present paper offers a review of recent developments in the CHARMM additive force field, one of the most popular biomolecular force fields. Additionally, we present a detailed discussion of the CHARMM Drude polarizable force field, anticipating a growth in the importance and utilization of polarizable force fields in the near future. Throughout the discussion emphasis is placed on the force fields' parametrization philosophy and methodology.
Major conclusions: Recent improvements in the CHARMM additive force field are mostly related to newly found weaknesses in the previous generation of additive force fields. Beyond the additive approximation is the newly available CHARMM Drude polarizable force field, which allows for MD simulations of up to 1 μs on proteins, DNA, lipids and carbohydrates.
General significance: Addressing the limitations ensures the reliability of the new CHARMM36 additive force field for the types of calculations that are presently coming into routine computational reach while the availability of the Drude polarizable force fields offers an inherently more accurate model of the underlying physical forces driving macromolecular structures and dynamics. This article is part of a Special Issue entitled "Recent developments of molecular dynamics".

1. Introduction

The scientific utility of computer simulations of physical reality at an atomistic level of detail has been compelling since the beginning of the information age. A sufficiently accurate simulation would be the ultimate microscope; among other things, it would make it possible to study proteins and other biophysically and medicinally relevant nanoscale structures in their native environment at femtosecond temporal and atomic spatial resolution. At the molecular level, reality is dominated by Quantum Mechanics (QM). Approximations of the Schrodinger Equation have been implemented in QM software packages, making it possible to model and simulate molecular reality at a high level of accuracy. However, such methods exhibit extremely poor computational scaling, and the prospect of applying them at biologically relevant system sizes and time scales remains elusive, prompting the need for more simplified models. This need is met by Molecular Mechanics (MM), which approximates atomic-scale reality using classical mechanics. Since no approximation is perfect, different approximations have been developed with different strengths and weaknesses; such approximations are called "force fields" [1].

Force fields consist of two parts. The first part is the potential energy function, which expresses the energy of the system as an analytical and easily differentiable function of the coordinates of the particles (atoms in the case of all-atom force fields) in the system. To achieve acceptable accuracies, these potential energy functions are highly parametric, giving rise to the second part: the parameter set. The combination of an appropriate parameter set and potential energy function (i.e. the force field) makes it possible to calculate the energy of and forces on a physical system as a function of its geometry, allowing a direct simulation of the thermal motion of the system through numerical integration of Newton's equations of motion. Such simulations are called Molecular Dynamics (MD), which is the subject of this special issue. With the computer power that is currently readily available to most research labs, MD simulations on biologically relevant systems involving hundreds of thousands of atoms can be performed for hundreds of nanoseconds and, with dedicated supercomputers, millisecond time scales [2] and micrometer-scale structures such as virus particles [3] are within reach.

The overall accuracy of the results obtained by any given MD simulation is determined by the sampling of the relevant conformations and

☆ This article is part of a Special Issue entitled "Recent developments of molecular dynamics".

* Corresponding author.
E-mail address: alex@outerbanks.umaryland.edu (A.D. MacKerell).

http://dx.doi.org/10.1016/j.bbagen.2014.08.004
0304-4165/

the force field. The sampling aspect is related to the aforementioned simulation time scale or to the utilization of enhanced sampling technologies, as discussed elsewhere in this special issue. For simulations of biological systems, a number of specialized empirical force fields are available [4–7]. Among these force fields, CHARMM has been one of the pioneers, and in its present-day incarnation, is one of the most well-established force fields for MD studies of biomolecular systems. In this paper, we will focus on the different components of the CHARMM36 all-atom additive force field, and discuss the recent advancement of difference components of the model. This includes the protein [8] nucleic acid, [9,10], lipid [11], carbohydrate [12] and general organic molecule [13] parameter sets. In addition, we will present the recently released CHARMM Drude polarizable force field, which is representative of the next generation of MM-based simulation methods. Additional background information on both additive and polarizable force fields may be found in a number of recently published reviews [14–17].

2. Potential energy functions

2.1. The CHARMM Class I additive potential energy function

The CHARMM36 additive force field uses the Class I additive potential energy function, the different terms of which are given by Eq. (1).

$$\begin{aligned}&\text{Bonded terms}\\ &E_{bonded}=\sum_{bonds}K_b(b-b_0)^2+\sum_{angles}K_\theta(\theta-\theta_0)^2+\sum_{\substack{improper\\ dihedrals}}K_\varphi(\varphi-\varphi_0)^2\\ &\qquad+\sum_{dihedrals}\sum_{n=1}^{6}K_{\phi,n}(1+\cos(n\phi-\delta_n))\\ &\text{Nonbonded terms}\\ &E_{nonbonded}=\sum_{\substack{nonbonded\\ pairs\,i,\,j}}\frac{q_iq_j}{4\pi D\left\|\mathbf{r}_i-\mathbf{r}_j\right\|}+\sum_{\substack{nonbonded\\ pairs\,i,\,j}}\varepsilon_{ij}\left[\left(\frac{R_{\min,ij}}{\left\|\mathbf{r}_i-\mathbf{r}_j\right\|}\right)^{12}-2\left(\frac{R_{\min,ij}}{\left\|\mathbf{r}_i-\mathbf{r}_j\right\|}\right)^{6}\right].\end{aligned}\tag{1}$$

The bonded part of the potential energy function (sometimes less precisely referred to as the intramolecular or internal terms) consists of terms that correspond to connectivity patterns of atoms within a molecule; specifically, b, θ, φ and ϕ respectively represent the bond lengths, valence angles, improper dihedral angles and dihedral angles that are determined from the molecular geometry. For the first three terms, b_0, θ_0, and φ_0 are the corresponding bond, angle and improper equilibrium values and K_b, K_θ and K_φ are the respective force constants. As for the dihedral term, $K_{\phi,n}$ is the amplitude and δ_n is the phase for each multiplicity n. The nonbonded portion consists of an electrostatic term, where D is the dielectric constant (which needs to be set at its vacuum value ε_0 in order for the CHARMM force field to behave correctly) and q_i and q_j are the respective partial atomic charges on atoms i and j, and a van der Waals (vdW) term, which is treated by the Lennard–Jones (LJ) 6–12 potential in which ε_{ij} is the well depth and $R_{\min,ij}$ is the radius. In both of the nonbonded terms, $\|\mathbf{r}_i-\mathbf{r}_j\|$ is the distance between atoms i and j.

In the CHARMM force field, the bonded part of the potential energy function is augmented with two extra terms. The Urey–Bradley term (Eq. (2)) is a harmonic potential between the terminal atoms (1,3) that define a valence angle, providing modest improvements in the vibrational modes of model compounds. The CMAP ("correction map") term is a cross term between two adjacent dihedrals, and is currently only applied to the backbone ϕ,ψ dihedrals in proteins, where it improves the conformational properties in general and secondary structure propensities in particular. It is implemented as a 2D grid of energy corrections in (ϕ,ψ) space; energies in-between grid points are obtained by bicubic spline interpolation [18].

$$E_{UB}=\sum_{angles\,1,2,3}K_{UB}\left(r_{1,3}-r_{1,3;0}\right)^2.\tag{2}$$

It was recognized early in the history of MM that, compared to physical reality, the above Class I potential energy function suffered from lack of anharmonicity and cross terms, which was remedied by the introduction of Class II potential energy functions [19–25]. Similarly, the r_{ij}^{-12} part of the LJ potential is motivated by computational ease rather than physical relevance, and more accurate alternatives have been proposed. However, the determination of parameter sets for these more refined potential energy functions has suffered from lack of suitable target data, both quantitatively (the larger number of parameters greatly exacerbates the underdetermined nature of the parameter optimization process) and qualitatively (specific target data is needed to parametrize e.g. a Buckingham-type repulsive potential, and this is not always easy to obtain). More successful were the (largely parallel) efforts to develop Class I force field parameters that reproduce experimental bulk phase properties, and at the time of writing, the resulting Class I condensed phase force fields are by far the most popular for bulk phase simulations in general and biomolecular simulations in particular. At the same time, it has become clear that the fixed atom-centered point charges in these force fields needed to be optimized for a specific dielectric environment, and transfer poorly to other dielectric environments. Indeed, while this additive approximation is computational efficient and useful because the energy of a system can be expressed as the sum of the energies of the components plus the *pairwise* interaction energies for all permutations of *two* components, it is particularly problematic when simulating systems that contain regions with different dielectric properties, such as lipid bilayers in water. Also, as discussed in more detail below, because dihedral parameters around rotatable bonds are essentially correction terms for the 1–4 (i.e. terminal atoms of a dihedral angle) and longer-range intramolecular nonbonded interactions, they are made more transferanble by making the treatment of the electrostatic interactions more accurate. For these reasons, there is a growing consensus that the lack of polarizability is the most important factor limiting the accuracy of the current generation of biomolecular force fields.

2.2. Inclusion of polarizability through the Classical Drude Oscillator model

There are three common schemes used for the introduction of polarization into an energy function [17,26–28]. These are the fluctuating charge model, the polarizable dipole (a.k.a. induced dipole) model, and the classical Drude oscillator (a.k.a. Shell or charge-on-spring) model. We limit this discussion to the latter, as used in the CHARMM Drude polarizable force field because of its simplicity and relative ease of implementation into existing MD engines. In the Drude oscillator model, a virtual particle ("Drude particle") with charge $q_{D,i}$ is connected to the core of each polarizable atom i by means of a harmonic potential ("spring") with force constant k_D. The extra nonbonded energy term added to the potential energy function by the introduction of the Drude particles is given by

$$E_{Drude}=\frac{1}{4\pi D}\left(\sum_{i<j}\frac{q_{D,i}q_j}{\left\|\mathbf{r}_{D,i}-\mathbf{r}_j\right\|}+\sum_{i<j}\frac{q_{D,i}q_{D,j}}{\left\|\mathbf{r}_{D,i}-\mathbf{r}_{D,j}\right\|}\right)+\frac{1}{2}\sum_{\substack{polarizable\\ atoms\,i}}k_D\left\|\mathbf{r}_{D,i}-\mathbf{r}_i\right\|^2.\tag{3}$$

Subject to the spring's restraining force, the position $\mathbf{r}_{D,i}$ of the Drude particle (which carries a negative charge in the CHARMM Drude model) is allowed to adapt freely to the environmental electrostatic potential, thereby representing the deformation of the atom's electron cloud in response to the electrostatic environment. The corresponding atomic polarizability α_i of atom i is given by $\alpha_i = q_{D,i}^2/k_D$. To limit the proliferation of parameters and the underdetermined nature of the parametrization, k_D is set to a constant value of 500 kcal/(mol Å^2) for all atoms in the model, as further elaborated in the section "Drude parametrization" below.

The conceptually simplest way to perform an MD simulation with Drude oscillators (or indeed, any polarizable model) is to allow the induced polarization to fully relax at every time step. As this full relaxation involves interactions between Drude particles (second term in Eq. (3)), it can only be solved in an iterative self-consistent fashion by performing an energy minimization of the Drude particles; this is called an "SCF calculation" by analogy to QM methods. Just like in the QM field, performing an SCF at every step (Born–Oppenheimer MD) is computationally expensive. To overcome this, analogous to the Car–Parrinello method [29] used in QM MD, Drude simulations are commonly performed using an Extended Lagrange method, where the Drude particles are given a small mass, typically 0.4 amu, that is taken from the parent atom. The Drude particles are then made part of the equations of motions, the integration of which is modified such that the force associated with the Drude particles never becomes large, allowing the use of a time step (typically 1 fs) that is only modestly smaller than without them. This is most often accomplished by using two Nosé–Hoover thermostats: one to keep the real atoms at room temperature and a second one to keep the temperature associated with the motion of the Drude particles relative to their nuclei as low as possible, yet high enough for the induced polarization to react to the motion of the atoms [30–32]. In the extended Lagrangian method developed by Lamoureux and Roux, the "real atom" thermostat is actually attached to the center of mass of the parent atom and Drude particle due to a part of the atomic mass being shifted to the latter [31].

A particularly common problem with polarizable MD simulations is "polarization catastrophe", where an unphysically high polarization is mutually reinforced between atoms that are too close in space. It can be shown [33] that this metastable behavior is inherent to polarizable models in which the induced dipole moment is a linear function of the electric field. In the Drude model, polarization catastrophe manifests itself by the Drude particle of one atom being locked into the potential well of an adjacent atom. Another formally related issue is that a subsystem of 2 nearby atoms is more polarizable along the axis formed by the 2 atoms than perpendicular to it, to an unphysical extent. To suppress these undesirable effects, additional factors were introduced in Eq. (3), as explained in the section "Extensions of the potential energy function" below.

Two additional extensions of the Drude model were implemented to improve the treatment of nonbonded interactions as a function of orientation of hydrogen bond acceptors with the environment: virtual particles representative of lone pairs and anisotropic atomic polarizability. Lone pairs, which typically carry the charge on a hydrogen bond acceptor, approximate the presence of a static dipole and quadrupole moment on the atom. Anisotropic atom polarizabilty, which is an extension of Eq. (3) that describes an isotropic polarizability, accounts for the polarizability of physical atoms in molecules having significant anisotropic character. The CHARMM Drude polarizable force field contains parameters to explicitly capture this anisotropy, requiring more refinements to Eq. (3). These two terms are described in more detail below.

3. The CHARMM36 additive force field

3.1. Historical perspective

While the precursors to the CHARMM force fields historically have been among the earliest force fields used for biomolecular simulations [34], playing an important role in the birth of the field of biomolecular MD, the potentials used at that time differed significantly from present-day force fields in parametrization philosophy. Also, they were united-atom potentials, where nonpolar hydrogen atoms were not represented by explicit particles, being integrated into their parent atom instead. The latest iteration of the CHARMM united-atom force field was CHARMM19 [35], which was released in 1985 and is no longer under further development. Instead, parametrization efforts in the additive CHARMM force fields have focused on the all-atom model, where all atoms, including hydrogens, are treated explicitly. The first of these was the CHARMM22 force field in 1992, whose protein parameter set became a standard for years to come [36], and which also included early nucleic acid [37] and lipid [38] parameter sets. Support for a number of common ions followed shortly afterwards. A significant update followed with the CHARMM27 force field approximately 2000, featuring optimized nucleic acid [39,40] and lipid [41] parameter sets, along with the introduction of a number of relevant ions. The next improvement came in 2003 with the addition of the 2D dihedral correction map (CMAP) term to the CHARMM22 protein parameter set [42,18]. The result, named CHARM22/CMAP, represented a significant advance in the accuracy of protein backbone behavior. This was followed in 2005 with a lipid parameter set with improved acyl torsional potential referred to as CHARMM27r, [43] in 2008 with the release of the CHARMM35 carbohydrate parameter set, and in 2009 with the first release of the CHARMM General Force Field (CGenFF) [13]. While at this point, a relatively complete coverage of molecules commonly encountered in computational biophysics and computer-aided drug design projects was achieved, shortcomings in these force fields started to be brought to light by the availability of results from simulations on longer time scales [44] and other factors. The remainder of this section will mainly discuss these shortcomings and their remedies as implemented in the CHARMM36 force field, which was largely released in 2012. Apart from the physical improvements, it also features enhancements of a more technical nature, such as modularization of the different parameter sets, and elimination of the confusion in nomenclature that had arisen from different parameter sets (and improvements thereof) being released at different times.

3.2. Parametrization philosophy and methodology

The applicability and quality of any force field parameter set are a reflection of the emphases and thoroughness during its parametrization. Therefore, any discussion of the recent improvements in the different CHARMM additive parameter sets can only be understood in light of the parameterization philosophy and methodology. Inspired by the pioneering work of Jorgensen et al., the LJ parameters in the CHARMM additive force field are largely based on reproducing densities and heats of vaporization of bulk liquids, which ensured relevance for condensed-phase simulations [45–48]. Parameters obtained in this fashion were found to be fairly transferable between atom types of the same element and hybridization state, underlining their physical relevance [13]. For cases where the appropriate experimental data was not available, e.g. the model compound had a net charge or a decomposition temperature that was lower than its melting point, other sources of experimental target data were considered, such as crystal lattice parameters, heats of sublimation, and as available computing power grew, free energies of solvation [49]. Also, from CHARMM27 onwards, the use of experimental target data was combined with QM noble gas interaction data, mainly to obtain additional information regarding the relative LJ radii and well depths [47,48]. The partial charges were optimized based on the QM interaction energies between the model compound and a water molecule, placed as monohydrates at many different interaction sites. The distance between the water molecule and the model compound is optimized for each interaction site, and the MM charges on the latter are optimized to reproduce the resulting QM interaction energy, and to a lesser extent the distance. For this purpose, a scaling factor was applied to the vacuum QM interaction energy to account for the higher average polarization in aqueous medium. Likewise, systematic differences in the interaction distances were expected and observed using the above bulk phase LJ parameters, with the amount of deviation strongly depending on the type of interaction site [4,13,50]. As for the bonded parameters, those for the "hard" degrees of freedom, i.e. bonds, angles and (where applicable) Urey–Bradley and improper dihedrals were optimized targeting QM MP2 vibrational

modes for the force constants and a combination of MP2 optimized geometries and data obtained from surveys of the Cambridge Structural Database (CSD) [51] for the reference values [18,36]. Dihedral parameters were derived from QM Potential Energy Scans (PES) in an increasingly automated fashion [39,43,52,53]. Last but not least, the parameters from the small molecule optimization described above were assembled into force field representations of larger compounds with a more direct relevance for biomacromolecules (e.g. short peptides, small proteins, nucleotides, short DNA strands,…) that had been extensively studied experimentally. Simulations were performed to reproduce said experimental studies, and the parameters (especially the dihedrals) were fine-tuned in order for the simulation results to better match experiment.

3.3. Protein parameter set [8]

Although the CHARMM22/CMAP protein force field has proven its predictivity in studying protein dynamics, protein–protein interactions and pharmacological applications through many years of widespread usage, the availability of longer simulation time scales brought some deficiencies to light. Specifically, when studying the folding of the human Pin1 WW β-domain through direct unbiased simulation starting from the unfolded state, a misfolded structure consisting of helices results, the force field representation of which was subsequently confirmed by free energy calculations to be thermodynamically more stable than the native state [44]. Together with the knowledge that the chicken villin headpiece HP-35 α-subdomain folds correctly [54], this suggested that the α-helix propensity is too high in the CHARMM22/CMAP force field. To remedy this, the relative energies of the extended and helical regions of the CMAP potential for all amino acids except Glycine and Proline were adjusted [8] to reproduce NMR J-coupling data on the Ala_5 peptide [55] as well as chemical shift data reflecting helix formation on the 15-residue Ac-$(Ala\text{-}Ala\text{-}Gln\text{-}Ala\text{-}Ala)_3$-$NH_2$ peptide [56]. Since no such extensive experimental data sets were found for Glycine and Proline, which are known to have substantially different Ramachandran plots, the CMAPs on these two amino acids were rebuilt based on vacuum RIMP2/CBS//MP2/aug-cc-pVDZ calculations.

In parallel, the opportunity of refining the backbone potential was exploited to perform a systematic reparametrization of side chain dihedrals [8], which previously contained a lot of "default" values that were transferred into CHARMM22 from an earlier alkane parameter optimization project. To this end, a set of model compounds was created by capping each of the 20 standard amino acids except Glycine, Alanine and Proline with an acetyl N-terminus and a methylamine C-terminus (i.e. dipeptides). For Histidine, the two neutral tautomers as well as the positive protonation state were considered, for a total of 19 model compounds. For each of these compounds, three (χ_1,χ_2) (where applicable) 2D relaxed PES were performed, with the backbone dihedrals respectively in the α_R, β or α_L conformations [57]. All these calculations were done at the RIMP2/cc-pVTZ//MP2/6-31G* level of theory. Then, all the dihedral parameters corresponding to the χ_1 and χ_2 rotatable bonds were fit to these QM results. The quality of the initial fit was mediocre because many amino acids share the same χ_1 parameters, including the ones with short charged side chains that have very specific energy profiles owing to electrostatic interactions with the backbone. This was remedied by introducing unique atom types for the C^β atom of Aspartate, Glutamate and protonated Histidine; it should be noted that this makes the CHARMM36 atom typing scheme incompatible with older CHARMM versions. The ensuing optimized parameters were the ones used for the aforementioned MD simulations, and the χ_1 and χ_2 dihedral probability distributions resulting from these simulations were compared to crystal data obtained from a survey of the Protein Data Bank (PDB) [58]. Where significant discrepancies existed, the parameters were tweaked accordingly. This same adjustment procedure was repeated targeting NMR J-coupling data for ubiquitin and GB1 in 8 M urea, which only slightly worsened the agreement with the QM and crystal target data for all amino acids except Glutamate. The final side chain and backbone parameters yielded improved treatment of conformational sampling resulting in more realistic secondary structure propensities, folding and sampling of disordered domains, as well as improvements in the side chain energetics that are anticipated to result in more accurate sampling of χ rotamers.

3.4. Nucleic acid parameter set [9,10]

Application of the CHARMM27 nucleic acid model, including the availability of longer simulations, identified an overestimated presence of local base pair opening in RNA, particularly for GC-rich regions, which generally do not open on the sub-microsecond time scales in experiment [59]. As this problem was completely absent in the CHARMM27 DNA representation, which uses the same parameters except for the sugar dihedrals, and differs from RNA only in the presence of the 2′-hydroxyl group, suspicion arose that the discrepancy was a result of incorrect sampling of this group. Detailed analysis of simulations indicated that the model primarily samples the O3′ orientation of the 2′OH group (i.e. hydroxyl proton directed towards the phosphodiester backbone) versus the base orientation (i.e. hydroxyl proton directed towards the minor groove), which NMR experiments indicated to be the primary orientation in duplex RNA. Therefore, QM relaxed PES were performed on 6 RNA model compounds at the RIMP2/cc-pVTZ//MP2/6–31+G(d) level of theory. Although the corresponding MM energy profiles were in qualitative agreement, a quantitative discrepancy existed in the 0–150° region, with the CHARMM27 relative energy consistently higher than the QM relative energy. As this 0–150° region corresponds to the base orientation, it was hypothesized that this overestimation in CHARMM27 caused the undersampling of that orientation, leading to base opening. These results were consistent with a combined QM/bioinformatics study indicating the orientation of the 2′OH to significantly impact the conformational properties of the RNA phosphodiester backbone [60]. This discrepancy was remedied to various extents for the 6 model compounds, resulting in 5 proposed parameter sets with different 2′OH dihedral parameters. Each of these parameter sets was validated thoroughly by simulating 14 diverse RNA structures in explicit water, and comparing the results to crystallographic and NMR data, with a special emphasis on J-coupling constants. Also studied were water probability distributions, which were compared to survey results of crystal water molecules in 22 high-resolution RNA crystal structures. As a final test, potentials of mean force (PMF) were calculated for the unfolding of two RNA hairpin structures. Based on all these results, most of which were in fair agreement with experiment for most of the new proposed parameter sets, 1 of the 5 proposed parameter sets was chosen to become part of the CHARMM36 force field.

Another discrepancy that came to light was that the CHARMM27 force field significantly underestimated the relative population of the BII substate of the canonical B form of DNA relative to the BI state [61–63]. To cure this, the dihedral parameters associated with the ε and ζ torsions were targeted. Additionally, it was necessary to reoptimize the C2′–C3′–C4′–O4′ torsion which influences the relative energies of the north and south sugar ring puckers, in order to retain a correct equilibrium with the A form of DNA. The methodology was similar to the 2′OH reparametrization discussed above, starting from QM 1D and 2D PES on a set of relevant model compounds. The dihedral parameters were subsequently subjected to 5 iterations of manual tweaking, targeting not only the QM energy profiles but also relative populations of the A, BI and BII forms in a "training set" of three oligonucleotides that were simulated each time in explicit water: GTAC2, BDJ025 and the EcoRI dodecamer, the sequences of which respectively favor A, B and B states. Special attention was paid to the BI/BII equilibrium in the latter two species. The final parameter set was then validated by simulating 8 diverse DNA structures in explicit water. The resulting agreement with experiment indicated that the parameters were properly optimized and ready to be released as part of CHARMM36.

3.5. Lipid parameter set [11]

In contrast to the issues in the protein and DNA parameters sets that came to light after their release, the most important shortcoming in the lipid parameter set was well-known at the time CHARMM27 and CHARMM27r were released [43,64]. Namely, the surface area per lipid in CHARMM27(r) bilayers needed to be maintained at its experimental value by use of an NPAT ensemble, where the "A" stands for constant area. In the absence of this area constraint, unphysical surface tension would cause the membrane to shrink to near gel-phase. While relevant results could be obtained by simulating CHARMM27(r) lipid bilayers in the recommended NPAT ensemble, this somewhat restricted the general applicability of the force field, and the associated high surface tension raised concerns about some results. Unlike the issues in the protein and DNA representation that could be addressed by subtle tweaks in the dihedral potential, the shortcomings in the lipid potential were the result of fundamental imbalances in the nonbonded (mainly electrostatic) forces, and the solution was a thorough reparametrization of the lipid parameter set. This effort mostly followed the procedure outlined in the section "Parametrization philosophy and methodology" above, with the addition of free energy of solvation calculations early in the process, specifically during the determination of the LJ parameters and charges on the head group model compounds methyl acetate and dimethyl phosphate. Interestingly, it was found that a correct area per lipid could only be achieved by overestimating the solvation free energy of the methyl acetate model compound, more so than in the corresponding CHARMM27(r) model. This is in line with previous work [65], and could possibly be explained by the redistribution of partial charges in the glyceride chemical environment.

The second flaw that was addressed in CHARMM36 was CHARMM27r's inability to reproduce the experimental deuterium order parameters in the glycerol and upper aliphatic chain regions, pointing to subtle defects in the structure at the lipid–water interface [43]. This problem was approached by reoptimizing the dihedral parameters targeting "Hybrid Methods for Interaction Energies" (HM-IE) [66] extrapolated CCSD(T)/cc-pVQZ//MP2/cc-pVDZ conformational energies for some torsions and directly calculated RIMP2/cc-pVQZ//MP2/cc-pVDZ for the torsions on larger model compounds. Subsequently, the dihedral parameters were refined to closely reproduce the aforementioned deuterium order parameters in MD simulations of DPPC bilayers, making the lipid parameter set in CHARMM36 the first one to do so. Just like for the protein and nucleic acid parameter sets, these experiment-based adjustments did not strongly compromise the agreement with the QM profiles. DPPC validation studies showed the model to reproduce deuterium order parameters and spin–lattice relaxation rates from ^{13}C NMR, which are related to the dynamic behavior, and experimental values related to the overall structure of the bilayer, specifically form factors and electron density profiles derived from X-ray diffraction experiments. Further validation of the final CHARMM36 model included calculations on the lipids DMPC, DMPC, DOPC, POPC and POPE, showing satisfactory agreement with experiment. Known limitations in the CHARMM36 lipid parameters are (1) a discrepancy with the experimental dipole potential drop across the DPPC bilayer, which might be due to the lack of polarizability forcing cancelation of errors into the model [67], and (2) a discrepancy between the surface tensions of bilayers and monolayers that is dependent on the inclusion of long-range LJ interactions in the simulation. The practical consequence is that, while CHARMM36 bilayers perform better *without* a correction term for long-range LJ interactions, monolayers surface tensions are only reasonable *with* this correction term, the latter observation opposite to CHARMM27r.

Recently, CHARMM36 parameters for the lipid sphinomyelin were published [68]. As with the remaining lipids, the optimization targeted a range of QM PES for model compounds followed by iterative optimization to reproduce selected experimental data. Special attention was paid to optimization of the nonbonded parameters in the ceramide head group. The resulting model reproduces electron density profiles, NMR J-couplings and order parameters and other observables. Interestingly, application of the model to pure *N*-palmitoyl sphingomyelin bilayers identified the presence of positive curvature, which may be considered a prediction of the model.

3.6. Carbohydrate parameter set [12,69–72]

In contrast to the parameter sets discussed previously, the carbohydrate parameter set was only released recently, representing the last class of biomolecules to be covered by the CHARMM additive force field. Consequently, its initial parametrization incorporated the new methodological elements and computationally expensive simulations that allowed the improvements in the other biomolecular parameter sets outlined above. Because basic carbohydrates are chemically simpler than the biomolecules discussed above, containing only aliphatic carbons and alcohol, ether, aldehyde and ketone groups, few new LJ parameters were required. Also, although their electrostatic interactions are nontrivial, the requirement of standardized transferable charges limited the charge optimization effort. Therefore, a large part of the parametrization of the carbohydrate force field consisted of optimizing bonded parameters, more specifically dihedrals as carbohydrates have rich conformational behavior. Target data for the initial optimization were MP2/cc-pVTZ//MP2/6-31G(d) potential energy surfaces. However, for rings, simple relaxed PES are inappropriate; in particular, the conformational flexibility of 5-membered rings is best described by a complicated pseudorotational profile rather than simple torsions [70]. Therefore, the PES for the cyclic model compounds consisted of a number of preselected relevant ring conformations. Also of note is the inclusion of energy differences associated with anomeric equilibria in the dihedral fitting; this makes it possible to use the CHARMM36 force field to perform e.g. free energy perturbations between anomers. For the experiment-driven adjustment stage of the optimization, NMR J-coupling constants played a prominent role, supplemented with NOE data and simulations on 18 carbohydrate crystals, the structures of which were obtained from the CSD. Apart from basic carbohydrates, a few common derivatives such as xylose, fucose, N-acetylglucosamine (GlcNAc), N-acetylgalactosamine (GlaNAc), glucuronic acid, iduronic acid, and N-acetylneuraminic acid(Neu5Ac) were parametrized in this fashion [12]. Validation MD studies included examining the conformational properties and stability of oligomeric hyaluronan, sialyl Lewis X and the acetamido group in GlcNAc, as well as 4 larger glycoprotein systems and the intermolecular interactions in a lectin–sucrose complex. Subsequently, the model has been applied to study glycopeptides, including the anti-proliferative factor [73], and validated against NMR data for di- and trisaccharides [74]; these efforts included targeting the 1–6 linkage, which led to additional optimization of associated dihedral parameters [75]. The presence of a broad carbohydrate model that is consistent with the remainder of the CHARMM36 force field is anticipated to facilitate studies of complex systems, such as glycolipids, glycoproteins and a range of carbohydrate-containing antibiotics [76].

While lipid bilayers consist of separate lipid molecules, and proteins and nucleic acids are chemically relatively simple linear polymers than can often be studied satisfactorily without covalent adducts, carbohydrates of biological interest are more often than not branched and covalently attached to other biomolecules such as a proteins. This makes the preparation of a carbohydrate-containing simulation system far less trivial than for the other biomolecules. Indeed, while work is still in progress to support a wider variety of glycosidic linkages to a wider variety of biomolecules, the technical difficulty of using the linkages that are already supported forms a significant limitation. To address this issue, the popular CHARMM-GUI web interface (www.charmm-gui.org) [77] for simulation input generation was expanded with a Glycan Reader module [78] that automatically detects sugar-like structures in an input PDB file, determines the correct types of sugars and glycosidic linkages based on the connectivity and stereochemistry, and generates

input files that can directly be run with the CHARMM [79] and NAMD [80] simulation programs.

3.7. CHARMM General Force Field (CGenFF) [13,81,82]

CGenFF represents a more recent extension of the CHARMM additive all-atom force field motivated by the need to represent drug-like molecules and cofactors with a force field that is physically compatible with the biomolecular parts of the force field [13]. The starting point of CGenFF was the collection of model compounds used for the biomolecular FF optimization, after resolution of the overlap between parameters that arose from applying uniform CGenFF atom types to different biomolecules (e.g. 5-membered aliphatic ring parameters from proline, the nucleic acid backbone and the furanoses in the carbohydrate parameter set). The range of chemical space was then systematically expanded following a strategy designed to maintain consistency with the biomolecules. However, the parametrization strategy differs in details from its biomolecular counterpart due to practical considerations. With the biomolecules, which represent a limited region of chemical space, achieving the subtle balance of forces to treat macromolecules correctly in the condensed phase requires the highly laborious in-depth parametrization discussed in the previous sections. In contrast, drug-like molecules cover a vast area of chemical space that cannot be studied in the same detail, requiring modification of the parametrization philosophy that partially sacrifices "depth" in favor of "width". Accordingly, the CGenFF parameterization is mostly based on gas phase QM target data, mimicking the initial stages of the optimization procedure of the biomolecules. Charges were generally derived from QM gas phase interactions with a water molecule at different positions, and bonded parameters were based on MP2/6–31G(d) optimized geometries, vibrational analyses and PES, as above. It should be noted that this methodology, including the scaling factors for the water interactions, has demonstrated its value in the past, as indicated by the fact that the tweaks in the biomolecular parameter sets discussed above did not typically involve the nonbonded parameters. Only the LJ parameters are based on experimental liquid density and heat of vaporization data, because using LJ parameters derived from bulk-phase experimental data is mandatory in order to obtain qualitatively correct bulk-phase behavior [45,46]. However, owing to the high transferability of LJ parameters, only a modest number of LJ parameters required optimization during the development of CGenFF, with only a small number of additional LJ parameters anticipated. Final validation of CGenFF consisted of calculating the pure solvent molecular volume and heat of vaporization of 111 and 95 compounds, respectively. The results were excellent, with an average signed deviation of 0.6% and an average absolute deviation of 2.1% for the molecular volume, and respective deviations of −0.3% and 7.0% for the heat of vaporization.

Because of the high chemical diversity of drug-like molecules, the initial release of CGenFF covered an important but limited core chemical space. Subsequent expansions of the chemical space included (among others) carbamates, thiocarbonyls, hydrazines, amidinium groups, acyclic enol ethers, 4-membered rings, epoxides, benzoic acid esters and aliphatic nitro groups. Additionally, a targeted effort was made to cover a variety of sulfonyl-containing functional groups, including sulfonamides, sulfones, sulfoxides, anionic sulfonates, sulfonic esters, sulfamates and neutral organic sulfonates [83]. As the coverage gradually expanded, regular updates were made available, causing CGenFF to adapt a separate release cycle from the CHARMM biomolecular parameter sets. It should be emphasized that this does not imply that it is a separate force field; instead, CGenFF can be considered an evolving part of CHARMM36 that is fully compatible with its biomolecular counterparts. In line with this, biopolymers in a CHARMM context should always be represented with the latter, not with CGenFF, which cannot be expected to accurately capture the subtle balance of forces achieved during the detailed optimization of the biomolecule parameters.

While CGenFF as a force field supports a wide and expanding range of chemical groups, a typical drug or cofactor almost always consists of different chemical groups linked together, and it is beyond current computational means to optimize parameters for all linkages involving all permutations of chemical groups. Therefore, in parallel with the ongoing expansion in coverage, an effort was started to automate the application of CGenFF to arbitrary organic molecules [81,82]. This effort culminated in the release of the CGenFF program (publicly available for non-profit use at cgenff.paramchem.org). Upon input of an arbitrary molecule, this program first recognizes certain features of atoms and bonds, such as valence, rings and aromaticity. The criteria for aromaticity were chosen so that they reproduce the assignment of aromatic atom types in the existing library of 477 model compounds, which in turn was motivated by reproducing the QM target data as accurately as possible while retaining chemical consistency. All these features were passed on to an atom typer that is based on a programmable decision tree, facilitating the implementation of the complex atom typing rules that arose from assigning atom types to reproduce target data, as opposed to the more customary approach of starting from a predefined atom typing palette. The next step is to determine which bonded parameters are required for the molecule that are not already present in the force fields, and assign values to these parameters by analogy. This process is based on a matrix that contains penalty scores for substituting any atom type with any atom type, reflecting the dissimilarity between any two atom types. Based on this concept, a Total Penalty Score (TPS) between two parameters can be defined as the sum of the penalty scores of the 2, 3 or 4 atom types that respectively define a bond, angle or dihedral parameter. Thus, for each required parameter, the parameter with the lowest TPS is returned to the user, along with the TPS itself, which is provided as a rough measure of the quality. Enhancements of this basic scheme include a multiplier for atom types that play a more important role in defining a parameter (i.e. both atoms defining a bond, the central atom in a valence angle or improper dihedral and the two central atoms in a dihedral), the use of a different ("nonbonded") penalty matrix for atoms whose role in the behavior of the parameter is predominantly nonbonded in nature (i.e. the outer atoms in angle, dihedral and improper parameters), and an extra penalty term that is based on bond properties rather than atom types and is added to the TPS. Finally, charges are assigned using a bond charge increment scheme, with an increment value associated with each existing bond parameter. As an extension, two increment values were associated with each angle parameter and three with each dihedral, allowing the method to capture longer-range effects. All increment values were optimized to reproduce the charge distributions on the aforementioned library of model compounds. When applied to an arbitrary organic molecule, for each required parameter, the bond charge increment associated with the existing parameter with the lowest TPS is applied, using the aforementioned nonbonded penalty matrix throughout for calculating the TPS. The TPS values of the chosen set of bond charge increments are also used to calculate a separate penalty score for each charge in the molecule, again yielding a measure for the quality of that charge. The availability of these penalty scores is a major advantage, allowing the user to quickly judge how well the functional groups and combinations thereof in their molecule of interest are supported. Since its release, the CGenFF program has seen substantial use, with more than 36,000 molecules having been uploaded by more than 1800 users.

4. The CHARMM Drude polarizable force field

4.1. Extensions of the potential energy function

The first generation Drude polarizable energy function was a simple extension of the additive energy function that treated polarization using the auxiliary Drude particles as described above. However, a number of issues were encountered during the development of the Drude force

field, requiring extensions of the energy function. Of primary concern was the potential for polarization catastrophe, as described in the section "Potential energy functions" above, due to the induced dipole moment being a linear function of the electric field. Accordingly, an obvious remedy would be to include quadratic and/or higher-order terms, i.e. hyperpolarizability, which would effectively suppress the problem in the polarizable dipole model [33]. During early development of the CHARMM Drude polarizable force field, this was implemented by adding higher-order (anharmonic) terms to the harmonic potential between the Drude particle and its parent atom. However, during force field development, it became clear that in long Drude simulations on ionic systems, there would always be cases of polarization catastrophe due to the electrostatic interaction between two opposite point charges (in this case, the Drude particle and the adjacent nucleus) going to infinity when the distance approaches zero, a problem that could not be overcome by the inclusion of hyperpolarizability. Therefore, the hyperpolarizability approach was supplemented with a "HardWall" modification to the integrator [84]. This treatment involves reversing and scaling down the radial component of the Drude particle's relative velocity vector whenever it moves beyond a selected distance from the parent atom's nucleus (typically 0.2 Å) while retaining an outward relative velocity. In this scenario, the parent atom's velocity is adjusted accordingly to satisfy preservation of momentum. As the scaling down implies an inelastic collision, there is a net loss of kinetic energy, which is absorbed by the dual thermostat method described above.

An additional issue in the Drude force field is interactions between Drude particles corresponding to 1–2 and 1–3 atom pairs. These interactions are included in the energy function to allow the 1–2 and 1–3 atomic dipoles to react to each other, thereby leading to improved treatment of molecular polarizabilities. However, to prevent these interactions from being represented incorrectly due to fundamental limitations of the model, those electrostatic terms are scaled by a screening function S_{ij} that depends on the distance r_{ij} between Drude particles i and j. Inspired by the work of Thole [33,85,86], the Thole-like screening function used in the CHARMM Drude model is [87]

$$S_{ij}\left(r_{ij}\right) = 1 - \left(1 + \frac{\left(a_i + a_j\right) r_{ij}}{2\left(\alpha_i \alpha_j\right)^{1/6}}\right) e^{-\frac{\left(a_i + a_j\right) r_{ij}}{\left(a_i a_j\right)^{1/6}}} \tag{4}$$

where a_i is called the "Thole damping parameter" on atom i and is defined for individual atoms in a similar fashion as the atomic polarizabilities α_i and partial charges q_i.

Two extensions of the Drude model to better capture the anisotropy of the charge distribution and of the polarizability are, respectively, lone pair particles and anisotropic polarizabilities. These are only applied on appropriate atoms, typically hydrogen bond acceptors, as mentioned above. Both are defined in a local reference frame centered on the atom in question, with its orientation defined by the positions of selected adjacent atoms. The lone pair particles are massless negative point charges, the position of which is fixed in the local reference frame. Anisotropy is introduced in the polarizability by replacing the force constant k_D with a force constant tensor $\mathbf{K}_D$, which is related to the polarizability tensor $\boldsymbol{\alpha}$ as $\mathbf{K}_D = -q_D^2\,\boldsymbol{\alpha}^{-1}$. To limit the proliferation of parameters, $\boldsymbol{\alpha}$ and $\mathbf{K}_D$ are chosen to be diagonal in the CHARMM Drude polarizable force field, and the harmonic potential connecting the Drude particle to its parent atom simply becomes $\mathbf{K}_{D,11}\,x_D^2 + \mathbf{K}_{D,22}\,y_D^2 + \mathbf{K}_{D,33}\,z_D^2$, with all quantities defined in the aforementioned local reference frame. As the relative contribution of the off-diagonal terms in atomic polarizabilities is generally small, this diagonal formulation is thought to be an acceptable approximation for the fully anisotropic polarizability. For the purpose of calculating the Thole-like screening factor, the isotropic polarizability $\alpha = 1/3\ \mathrm{Tr}(\boldsymbol{\alpha})$ is used.

4.2. Drude parametrization

Optimization of the electrostatic parameters, including the partial atomic charges q_i, polarizabilities α, and the Thole damping parameters a_i (Eq. (4)) is based on the reproduction of the B3LYP/aug-cc-pVDZ electrostatic potential (ESP) calculated at grid points located on concentric Connolly surfaces [88] around a chosen model compound, both in the absence and presence of small (typically $+0.5e$) perturbing point charges at various locations. The inclusion of the perturbed ESPs is required to simultaneously fit the charges, polarizabilities, and Thole terms, where the latter two terms are modeling the electronic response introduced by the ESP perturbations. We note that fitting of the polarizabilities α_i essentially involves fitting the charges on the Drude particles $q_{D,i}$ (which are subtracted from the atomic charges q_i) in Eq. (3) given that the force constant on the harmonic spring between the parent nuclei and the Drude particles is fixed. In addition, for the hydrogen bond acceptors it is necessary to fit the geometric parameters and charges of the lone pair particles and the polarizability tensors $\mathbf{K}_{D,i}$. Fitting of the lone pairs is often included as part of the ESP fitting, though it may be supplemented with QM model compound–water interactions. $\mathbf{K}_{D,i}$ values are typically fit against selected ESP points represented on arcs around the acceptor atom [89]. As indicated above, the lone pairs and anisotropic polarizabilities are included to more accurately treat interactions as a function of orientation, with the latter particularly of utility for interactions with ions.

To facilitate determination of the electrostatic model a number of issues need to be considered. Even though the number of target data points in the ESPs is high, there exist strong dependencies between them, and the final fitting problem is usually ill-conditioned. Therefore, it is crucial to restrain all the parameters involved in the fitting towards realistic reference or target values, and the merit function for the electrostatic fitting in function of n generic parameters p_i becomes:

$$\chi^2(p_1, K, p_n) = \sum_{\substack{grid \\ points\ j}}^{m} \left(V_j^{MM}(p_1, K, p_n) - V_j^{QM}\right)^2 + \sum_{parameters\ i}^{n} w_i\left(p_i - p_i^{ref}\right)^2 \tag{5}$$

where V_j is the electrostatic potential at grid point j, w_i is a weight factor associated with parameter i and p_i^{ref} is its aforementioned reference value. For the atomic charges, the best reference values were found to be their counterparts from the CHARMM additive force field. As mentioned above, the isotropic force constant k_D was set to a constant value of 500 kcal/(mol Å^2), allowing the target charges on the Drude particles $q_{D,i}^{ref}$ to be determined from Miller isotropic atomic polarizabilities [90] using the relationship $q_{D,i}^{ref} = \sqrt{\alpha_i^{Miller} k_D}$. The targets for the Thole damping parameters a_i^{ref} were set to the benzene value of 1.3 [91]. The targets for the geometric parameters of the lone pair particles were based on an Atoms in Molecules (AIM) analysis [92] of the electron density, while their charges were indirectly restrained by restraining the charge on the parent atom towards 0 so that, in the absence of other forces in the optimization, the latter's negative charge is preferentially channeled to its lone pairs (note that lone pairs were exclusively added to electronegative atoms). Where applicable, the anisotopic force constant tensors were obtained by multiplying the aforementioned isotropic k_D value by a diagonal anisotropy tensor $\mathbf{A}_{D,i}$, and, by setting of trace of $A_D = 3$, only two components, A_{11} and A_{22} need to be specified as $A_{33} = 3 - (A_{11} + A_{22})$. Finally, the atomic polarizabilities were scaled down a posteriori by a scaling factor that depends on the chemical nature of the atom, in order to account for the fact that molecules are less polarizable in the condensed phase than in the gas phase. This was typically performed empirically based on the reproduction of dielectric constants for pure solvents [26,93].

Like in the CHARMM additive force field, the LJ parameters in the CHARMM Drude polarizable force field were optimized using a labor-

and computer-intensive iterative procedure, targeting a wide variety of experimental bulk phase as well as QM target data. The range of target data for the CHARMM Drude polarizable force field was substantially wider than for its additive counterpart, comprising properties that partially depend on the polarizability such as the liquid phase dielectric constant, as well as the density, lattice geometry, enthalpy of vaporization and free energy of solvation, isothermal compressibility, self-diffusion coefficient, heat capacity and osmotic pressure, as available. Also, more systematic use was made of QM gas phase interaction data between the model compound and water molecules as well as rare gases [47,48]. Finally, optimization of bonded parameters was largely identical to the CHARMM additive force field, with a first stage that targets MP2 level gas phase conformational energies and vibrational analyses, and a second stage where empirical adjustments are made to correctly reproduce the behavior of biomacromolecules in aqueous solution. In this second stage, empirical tweaks to the electrostatic parameters were also made where appropriate. Compared to the additive force field, it was more difficult in the polarizable force field to perform these adjustments such that all the parameters remained balanced, but the required tweaks were also less drastic, owing to the polarizable force field's enhanced ability to simultaneously reproduce gas- and bulk phase properties.

4.3. Proteins

The procedure described in the previous section was applied to obtain parameters for the protein side chain model compounds and initial parameters for the backbone. However, additional special emphasis was put on the backbone electrostatics, which were further optimized targeting conformational energies, interactions with water, molecular dipole moments and polarizabilities of the Alanine dipeptide Ace-Ala-NMe at the QM level as well as experimental condensed phase data for extended polypeptides such as Ace-Ala_5-NMe. As in the CHARMM36 additive force field, CMAP terms were then introduced based on a 2D QM and MM PES in vacuum for Glycine, Proline, and Alanine, the latter representing all other amino acids. Subsequently, empirical corrections to the Alanine CMAP term were introduced to reproduce experimental bulk-phase behavior of peptides and small proteins. The side chain dihedral parameters were optimized initially targeting the gas phase QM-based side chain 1D and 2D PES mentioned in the section "Protein parameter set" above, followed by additional adjustment aimed at reproducing crystal-based experimental data on proteins using the model peptides Ace-Ala_4-X-Ala_4-NMe (where X is the amino acid of interest).

Validation of the Drude model has been performed on a number of peptides and proteins. In general the model yields root-mean-square differences (RMSD) with respect to crystal structures for C^{α} atoms similar to or larger than the CHARMM36 additive model on the 100 ns MD time scale and the agreement with NMR J-coupling for the $(Ala)_5$ peptide is $\chi^2 = 2.3$ versus 1.2 for the CHARMM36 additive FF, where χ^2 represents the weighted RMSD between experimental and calculated coupling constants. However, order parameters for the Drude model are in good agreement with experiment for ubiquitin, GB3 and cold shock protein A. More recently, simulations of ubiquitin and cold shock protein A were extended to 1 μs, yielding stable structures with RMSD values similar to those from CHARMM36 [94].

Notable from the Drude simulations are the magnitudes of the local dipole moments of peptide bonds and the side chains of Tryptophan residues. In both cases for multiple residues the dipoles are significantly larger than those in the additive CHARMM36 model, even though the latter were systematically overestimated to create an additive force field suitable for condensed phase simulations. In addition, the Drude model displayed significant variations in the dipole moments between different residues of the same type as well as within the same residue as a function of simulation time. For example, the dipole moments of the peptide bonds with CHARMM36 are ~3.8 D (~4.6 D for the alpha L conformation) versus values ranging from ~4.7 to 5.4 D with the Drude model.

An important result with the Drude model was its ability to significantly improve the treatment of the cooperativity of helix formation in the $(AAQAA)_3$ peptide versus the additive CHARMM36 model [95]. This cooperativity was associated with induction of peptide bond dipoles upon helix formation. The improved treatment by the Drude model is important as all additive force fields studied to date have been shown to significantly underestimate cooperativity of folding of this peptide [96]. Quantitative analyses included use of the helix-coil theory of Lifson–Roig [97] for which experimental helix nucleation and elongation parameters are known. The experimental elongation parameter w is 1.28 at 300 K, with the Drude value of 1.17 being in better agreement than the additive value of 1.03. In addition, the experimental nucleation parameter v of 0.04 is better reproduced by the Drude (0.11) as compared to the additive (0.17) force field. These results indicate that the Drude model produces a fundamentally more accurate treatment of the physical forces driving protein structure and function, which is anticipated to yield an improved understanding of conformational variability in peptides and proteins, including protein folding.

4.4. Nucleic acids

Development of a CHARMM Drude polarizable model for nucleic acids followed the above described approach of optimizing parameters for cyclic ethers [98,99], dimethylphosphate and the bases [100]. These were then combined in a polarizable model for DNA. Refinement of that model required significant adjustments of the phosphodiester backbone and glycosidic bond dihedral parameters, of base-sodium interactions and of selected electrostatic parameters in the model [101]. The resulting DNA force field yields RMSDs similar to the CHARMM36 force field for a number duplexes on the 100 ns timescale, improved agreement in NMR order parameters for the EcoRI dodecamer, and improved agreement with counterion condensation theory with respect to charge neutralization by NaCl. The polarizable model also satisfactorily treats the equilibrium between the A, BI and BII conformations of DNA [102,10]. Similar to the polarizable protein model, the dipole moments of the bases are significantly larger than in the additive CHARMM36 model.

Recently, the Drude DNA force field was applied to study base flipping in two sequences [103]. While only two systems were studied, the polarizable model gave significantly improved agreement with NMR data on the equilibrium between the open and closed (i.e. Watson–Crick base-paired) states, K_{op}. For AT opening with CHARMM36 a value of $K_{op} = 3.5 \times 10^{-8}$ was obtained versus a value of 4.3×10^{-6} with the Drude model, which is in significantly better agreement with the experimental value of $(2.00 \pm 0.04) \times 10^{-5}$. This improved agreement was shown to be largely due to the dipole moments of the bases increasing upon flipping out of the duplex (e.g. from ~3.5 to 4.9 D for the adenine base during A flipping), thereby leading to more favorable solvation that stabilized the open states. In addition, alterations of the dipole moments of water molecules adjacent to the bases occurred as the bases flipped out of both the minor and major grooves. These results further point to a fundamental difference in the physical forces driving macromolecular dynamics in the polarizable force field.

4.5. Lipids

Drude polarizable parameters have been presented for phosphatidylcholine lipids such as DPPC. A preliminary Drude model yielded significantly improved agreement for the membrane dipole potential of a DPPC monolayer–air interface [67]. Subsequently, an improved DPPC Drude model was presented [84]. That force field yields satisfactory agreement with the surface area/head group, X-ray scattering form factors, electron density profiles and NMR order parameters for a DPPC bilayer. While many of the results were similar to the highly optimized

additive CHARMM36 lipid force field, the Drude model gave improved results for monolayer surface tensions and the electrostatic potential across the DPPC bilayer.

4.6. Carbohydrates

Towards a polarizable carbohydrate Drude force field, models for polyalcohols and hexapyranoses have been presented [104,105]. For the polyalcohols, it was shown that the inclusion of polarization allowed for significantly improved reproduction of the conformational energies of ethylene glycol over the additive model due to the change in molecular dipole as a function of conformation. Another interesting observation was the ability of the Drude model to accurately reproduce the heat of vaporization of glycerol. Accurate reproduction of this value is not possible with the additive force field due to the need to intrinsically overpolarize the model with respect to vacuum in order for it to interact accurately with water, which has priority over gas phase properties. This leads to overstabilization of conformations in which the three vicinal hydroxyls are hydrogen bonded to each other, thereby artificially lowering the energy of the gas phase monomer leading to a systematic error in the heat of vaporization. As the polarizable model more accurately treats the electrostatic properties of both reference states in the heat of vaporization calculation (gas phase and pure solvent) simultaneously, this inherent problem is avoided. Optimization of the Drude hexapyranose model also showed the model to more accurately treat the molecular dipole moment as a function of conformation, leading to significant improvements in the reproduction of QM PES over the CHARMM36 additive force field. Notably, this was achieved for the 16 hexopyranose epimers and anomers using the same electrostatic model, indicating the improved transferability of those parameters over the additive model. Improvements were also seen in the reproduction of crystal structures of a number of monosaccharides. While significant extension of the polarizable carbohydrate force field to furanoses, disaccharides and non-hydroxyl moieties is underway, the results to date suggest that the Drude model will yield an inherently more accurate model of modeling carbohydrates over currently available additive models.

4.7. General organic molecules

Contrary to the CHARMM36 additive force field, one of the initial goals of the CHARMM Drude polarizable force field was to use the same basic atom typing palette for all types of biomolecules (proteins, nucleic acids, lipids and carbohydrates), with as few specific atom types for any single class of biomolecules as feasible without degrading the quality of that class's force field representation. The rationale for this was to avoid the error-prone and labor-intensive merging of parameters from different sets that was the starting point of CGenFF [13]. Also, choosing a well-defined atom typing palette early in the parametrization process may potentially lead to less complex atom typing rules [81]. To further improve the internal consistency of the CHARMM Drude polarizable force field as well as to facilitate the future creation of a "Drude General Force Field (DGenFF)", some of the biomolecular model compounds were parametrized in the framework of a larger series of congeneric molecules, not all of which have immediate biological applications. Examples of this are the alkanes [106] and the nitrogen-containing heteroaromatic compounds [107]. While the CHARMM Drude polarizable force field's coverage of chemical space is currently not wide enough to claim support for "general organic molecules", and priority is currently given to the biomolecular aspects of the force field, a solid basis for extending coverage of chemical space is being established by the updated parameterization philosophy discussed in this paragraph. Facilitating the development of the DGenFF is the development of an online parameter optimization utility referred to as GAAMP [108].

5. Summary

CHARMM36 not only features wide coverage of molecules commonly encountered in computational biophysics and computer-aided drug design, but also has been subject to recent meticulous efforts to ensure correct behavior at presently accessible simulation time scales. Additionally, a number of nontrivial system preparation tasks have been automated, most notably the attachment of oligosaccharides to polypeptide chains and the assignment of atom types, parameters and charges on general organic molecules. This makes CHARMM36 a versatile tool for biomolecular simulations, representing the state-of-the-art in additive force fields. Looking forward, the CHARMM Drude polarizable force field addresses the additive force fields' main fundamental weakness by including polarizability, and after a protracted period of parameter optimization and fine-tuning of the potential energy function, it has now reached a stage where it is ready for production calculations on proteins, DNA and selected lipids. As the computational cost of polarizable simulations is roughly a factor 2 higher than for additive ones when using the same MD integration time step [109], it is anticipated that the additive force field will continue seeing intensive use for years to come, and further automation efforts are underway on that front. Nevertheless, polarizable force fields are bound to become the standard for simulating systems that cannot be represented satisfactory with non-polarizable ones, and we speculate that from there, they will gradually take over the territory currently covered by additive force fields.

Acknowledgements

Financial support from the NIH (GM051501, GM070855 and GM072558), the NSF (CHE-0823198), the Waxman Foundation and the University of Maryland Computer-Aided Drug Design Center are acknowledged. Extensive collaboration and discussions with Prof. B. Roux and coworkers are greatly appreciated. Additionally, this research was supported in part by the NSF through TeraGrid [110] resources provided by NCSA and PSC.

References

[1] K. Vanommeslaeghe, O. Guvench, A.D. MacKerell Jr., Molecular mechanics, Curr. Pharm. Des. 20 (2014) 3281–3292.
[2] K. Lindorff-Larsen, S. Piana, R.O. Dror, D.E. Shaw, How fast-folding proteins fold, Science 334 (6055) (2011) 517–520.
[3] P.L. Freddolino, A.S. Arkhipov, S.B. Larson, A. McPherson, K. Schulten, Molecular dynamics simulations of the complete satellite tobacco mosaic virus, Structure 14 (2006) 437–449.
[4] A.D. MacKerell Jr., B. Brooks, C.L. Brooks III, L. Nilsson, B. Roux, Y. Won, M. Karplus, CHARMM: the energy function and its parameterization with an overview of the program, in: P.v.R. Schleyer, N.L. Allinger, T. Clark, J. Gasteiger, P.A. Kollman, H.F. Schaefer III, P.R. Schreiner (Eds.), Encyclopedia of Computational Chemistry, 1, John Wiley & Sons, Chichester, 1998, pp. 271–277.
[5] W.D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Merz Jr., D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell, P.A. Kollman, A second generation force field for the simulation of proteins, nucleic acids, and organic molecules, J. Am. Chem. Soc. 117 (19) (1995) 5179–5197.
[6] W.L. Jorgensen, D.S. Maxwell, J. Tirado-Rives, Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids, J. Am. Chem. Soc. 118 (1996) 11225–11236.
[7] C. Oostenbrink, A. Villa, A.E. Mark, W.F. Van Gunsteren, A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6, J. Comput. Chem. 25 (13) (2004) 1656–1676.
[8] R.B. Best, X. Zhu, J. Shim, P.E.M. Lopes, J. Mittal, M. Feig, A.D. MacKerell Jr., Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ and side-chain χ^1 and χ^2 dihedral angles, J. Chem. Theory Comput. 8 (9) (2012) 3257–3273.
[9] E.J. Denning, U.D. Priyakumar, L. Nilsson, A.D. MacKerell Jr., Impact of 2′-hydroxyl sampling on the conformational properties of RNA: update of the CHARMM all-atom additive force field for RNA, J. Comput. Chem. 32 (9) (2011) 1929–1943.
[10] K. Hart, N. Foloppe, C.M. Baker, E.J. Denning, L. Nilsson, A.D. MacKerell Jr., Optimization of the CHARMM additive force field for DNA: improved treatment of the BI/BII conformational equilibrium, J. Chem. Theory Comput. 8 (1) (2012) 348–362.
[11] J.B. Klauda, R.M. Venable, J.A. Freites, J.W. O'Connor, D.J. Tobias, C. Mondragon-Ramirez, I. Vorobyov, A.D. MacKerell Jr., R.W. Pastor, Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types, J. Phys. Chem. B 114 (23) (2010) 7830–7843.

[12] O. Guvench, S.S. Mallajosyula, E.P. Raman, E. Hatcher, K. Vanommeslaeghe, T.J. Foster, F.W. Jamison II, A.D. MacKerell Jr., CHARMM additive all-atom force field for carbohydrate derivatives and its utility in polysaccharide and carbohydrate-protein modeling, J. Chem. Theory Comput. 7 (10) (2011) 3162–3180.
[13] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, A.D. MacKerell Jr., CHARMM General Force Field (CGenFF): a force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields, J. Comput. Chem. 31 (4) (2010) 671–690.
[14] J.W. Ponder, D.A. Case, Force fields for protein simulations, Adv. Protein Chem. 66 (2003) 27–85.
[15] P. Cieplak, F.-Y. Dupradeau, Y. Duan, J. Wang, Polarization effects in molecular mechanical force fields, J. Phys. Condens. Matter 21 (2009) 333102.
[16] X. Zhu, P.E. Lopes, A.D. MacKerell Jr., Recent developments and applications of the CHARMM force fields, WIREs Comput. Mol. Sci. 2 (2012) 167–185, http://dx.doi.org/10.1002/wcms.74.
[17] Lopes, P.E.M.; Harder, E.; Roux, B.; MacKerell, A.D., Jr, Formalisms for the explicit inclusion of electronic polarizability in molecular modeling and dynamics studies, in Multi-scale Quantum Models for Biocatalysis: Modern Techniques and Applications, Volume editors Darrin York and Tai-Sung Lee; in the series, Challenges and Advances in Computational Chemistry and Physics, Vol. 7, Editor J. Leszczynski, p 219–218, 2009, ISBN: 978-1-4020-9955-7.
[18] A.D. MacKerell Jr., M. Feig, C.L. Brooks III, Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations, J. Comput. Chem. 25 (2004) 1400–1415.
[19] N.L. Allinger, Conformational analysis. 130. MM2. A hydrocarbon force field utilizing V1 and V2 torsional terms, J. Am. Chem. Soc. 99 (25) (1977) 8127–8134.
[20] N.L. Allinger, Y.H. Yuh, J.-H. Lii, Molecular mechanics. The MM3 force field for hydrocarbons. 1, J. Am. Chem. Soc. 111 (1989) 8551–8566.
[21] J.-H. Lii, N.L. Allinger, Molecular mechanics. The MM3 force field for hydrocarbons. 2. Vibrational frequencies and thermodynamics, J. Am. Chem. Soc. 111 (1989) 8566–8575.
[22] J.-H. Lii, N.L. Allinger, Molecular mechanics. The MM3 force field for hydrocarbons. 3. The van der Waals' potentials and crystal data for aliphatic and aromatic hydrocarbons, J. Am. Chem. Soc. 111 (1989) 8576–8582.
[23] N.L. Allinger, K.H. Chen, J.H. Lii, K.A. Durkin, Alcohols, ethers, carbohydrates, and related compounds. I. The MM4 force field for simple compounds, J. Comput. Chem. 24 (12) (2003) 1447–1472.
[24] M.J. Hwang, T.P. Stockfisch, A.T. Hagler, Derivation of class II force fields. 2. Derivation and characterization of a class II force field, CFF93, for the alkyl functional group and alkane molecules, J. Am. Chem. Soc. 116 (1994) 2515–2525.
[25] T.A. Halgren, Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94, J. Comput. Chem. 17 (5–6) (1996) 490–519.
[26] P.E.M. Lopes, B. Roux, A.D. Mackerell Jr., Molecular modeling and dynamics studies with explicit inclusion of electronic polarizability: theory and applications, Theor. Chem. Accounts 124 (2009) 11–28.
[27] S.W. Rick, S.J. Stuart, Potentials and algorithms for incorporating polarizability in computer simulations, Rev. Comput. Chem. 18 (2002) 89–146.
[28] F.J. Luque, F. Dehez, C. Chipot, M. Orozco, Polarization effects in molecular interactions, WIREs Comput. Mol. Sci. 1 (2011) 844–854.
[29] R. Car, M. Parrinello, Unified approach for molecular-dynamics and density-functional theory, Phys. Rev. Lett. 55 (1985) 2471–2474.
[30] M. Sprik, Computer-simulation of the dynamics of induced polarization fluctuations in water, J. Phys. Chem. 95 (6) (1991) 2283–2291.
[31] G. Lamoureux, B. Roux, Modeling induced polarization with classical Drude oscillators: theory and molecular dynamics simulation algorithm, J. Chem. Phys. 119 (6) (2003) 3025–3039.
[32] G. Lamoureux, A.D. MacKerell Jr., B. Roux, A simple polarizable model of water based on classical Drude oscillators, J. Chem. Phys. 119 (2003) 5185–5197.
[33] B.T. Thole, Molecular polarizabilities calculated with a modified dipole interaction, Chem. Phys. 59 (3) (1981) 341–350.
[34] J.A. McCammon, M. Karplus, The dynamic picture of protein structure, Acc. Chem. Res. 16 (1983) 187–193.
[35] E. Neria, S. Fischer, M. Karplus, Simulation of activation free energies in molecular systems, J. Chem. Phys. 105 (5) (1996) 1902–1921.
[36] A.D. MacKerell Jr., D. Bashford, M. Bellott, R.L. Dunbrack Jr., J. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, I.W.E. Reiher, B. Roux, M. Schlenkrich, J. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, M. Karplus, All-atom empirical potential for molecular modeling and dynamics studies of proteins, J. Phys. Chem. B 102 (1998) 3586–3616.
[37] A.D. MacKerell Jr., J. Wiórkiewicz-Kuczera, M. Karplus, An all-atom empirical energy function for the simulation of nucleic acids, J. Am. Chem. Soc. 117 (1995) 11946–11975.
[38] M. Schlenkrich, J. Brinkman, A.D. MacKerell, M. Karplus, An empirical potential energy function for phospholipids: criteria for parameter optimization and applications, in: K.M. Merz, B. Roux (Eds.), Membrane Structure and Dynamics, Birkhauser, Boston, 1996, pp. 31–81.
[39] N. Foloppe, A.D. MacKerell Jr., All-atom empirical force field for nucleic acids: 1) parameter optimization based on small molecule and condensed phase macromolecular target data, J. Comput. Chem. 21 (2000) 86–104.
[40] A.D. MacKerell Jr., N.K. Banavali, All-atom empirical force field for nucleic acids: 2) application to solution MD simulations of DNA, J. Comput. Chem. 21 (2000) 105–120.
[41] S.E. Feller, A.D. MacKerell Jr., An improved empirical potential energy function for molecular simulations of phospholipids, J. Phys. Chem. B 104 (2000) 7510–7515.
[42] A.D. MacKerell Jr., M. Feig, C.L. Brooks III, Accurate treatment of protein backbone conformational energetics in empirical force fields, J. Am. Chem. Soc. 126 (2004) 698–699.
[43] J.B. Klauda, B.R. Brooks, A.D. MacKerell Jr., R.M. Venable, R.W. Pastor, An ab initio study on the torsional surface of alkanes and its effect on molecular simulations of alkanes and a DPPC bilayer, J. Phys. Chem. B 109 (2005) 5300–5311.
[44] P.L. Freddolino, S. Park, B. Roux, K. Schulten, Force field bias in protein folding simulations, Biophys. J. 96 (2009) 3772–3780.
[45] W.L. Jorgensen, J.D. Madura, C.J. Swenson, Optimized intermolecular potential functions for liquid hydrocarbons, J. Am. Chem. Soc. 106 (1984) 6638–6646.
[46] W.L. Jorgensen, J. Tirado-Rives, The OPLS potential function for proteins. Energy minimizations for crystals of cyclic peptides and crambin, J. Am. Chem. Soc. 110 (1988) 1657–1666.
[47] D. Yin, A.D. MacKerell Jr., Combined ab initio/empirical approach for the optimization of Lennard–Jones parameters, J. Comput. Chem. 19 (3) (1998) 334–348.
[48] I.-J. Chen, D. Yin, A.D. MacKerell Jr., Combined ab initio/empirical approach for optimization of Lennard–Jones parameters for polar-neutral compounds, J. Comput. Chem. 23 (2) (2002) 199–213.
[49] Y. Deng, B. Roux, Hydration of amino acid side chains: non-polar and electrostatic contributions calculated from staged molecular dynamics free energy simulations with explicit water molecules, J. Phys. Chem. B 108 (42) (2004) 16567–16576.
[50] A.D. MacKerell Jr., M. Karplus, Importance of attractive van der Waals contributions in empirical energy function models for the heat of vaporization of polar liquids, J. Phys. Chem. 95 (26) (1991) 10559–10560.
[51] F.H. Allen, The Cambridge Structural Database: a quarter of a million crystal structures and rising, Acta Crystallogr. B 58 (2002) 380–388.
[52] O. Guvench, A.D. MacKerell Jr., Automated conformational energy fitting for force-field development, J. Mol. Model. 14 (8) (2008) 667–679.
[53] Vanommeslaeghe, K.; MacKerell, A.D., Jr; *manuscript in preparation.*
[54] P.L. Freddolino, K. Schulten, Common structural transitions in explicit-solvent simulations of Villin headpiece folding, Biophys. J. 97 (2009) 2338–2347.
[55] J. Graf, P.H. Nguyen, G. Stock, H. Schwalbe, Structure and dynamics of the homologous series of alanine. Peptides: a joint molecular dynamics/NMR study, J. Am. Chem. Soc. 129 (2007) 1179–1189.
[56] W. Shalongo, L. Dugad, E. Stellwagen, Distribution of helicity within the model peptide acetyl(AAQAA)$_3$amide, J. Am. Chem. Soc. 116 (1994) 8288–8293.
[57] X. Zhu, P.E.M. Lopes, J. Shim, A.D. MacKerell Jr., Intrinsic energy landscapes of amino acid side-chains, J. Chem. Inf. Model. 52 (2012) 1559–1572.
[58] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The Protein Data Bank, Nucleic Acids Res. 28 (2000) 235–242.
[59] Y. Pan, A.D. MacKerell Jr., Altered structural fluctuations in duplex RNA versus DNA: a conformational switch involving base pair opening, Nucleic Acids Res. 31 (2003) 7131–7140.
[60] E.J. Denning, A.D. MacKerell Jr., Intrinsic contribution of the 2′-hydroxyl to RNA conformational heterogeneity, J. Am. Chem. Soc. 134 (2012) 2800–2806.
[61] B. Heddi, N. Foloppe, N. Bouchemal, E. Hantz, B. Hartmann, Quantification of DNA BI/BII backbone states in solution. Implications for DNA overall structure and recognition, J. Am. Chem. Soc. 128 (2006) 9170–9177.
[62] B. Heddi, N. Foloppe, C. Oguey, B. Hartmann, Importance of accurate DNA structures in solution: the Jun–Fos model, J. Mol. Biol. 382 (2008) 956–970.
[63] K. Hart, L. Nilsson, Investigation of transcription factor Ndt80 affinity differences for wild type and mutant DNA: a molecular dynamics study, Proteins Struct. Funct. Bioinforma. 73 (2008) 325–337.
[64] S.E. Feller, K. Gawrisch, A.D. MacKerell Jr., Polyunsaturated fatty acids in lipid bilayers: intrinsic and environmental contributions to their unique physical properties, J. Am. Chem. Soc. 124 (2002) 318–326.
[65] J. Sonne, M.O. Jensen, F.Y. Hansen, L. Hemmingsen, G.H. Peters, Reparameterization of all-atom dipalmitoylphosphatidylcholine lipid parameters enables simulation of fluid bilayers at zero tension, Biophys. J. 92 (2007) 4157–4167.
[66] J.B. Klauda, S.L. Garrison, J. Jiang, G. Arora, S.I. Sandler, HM-IE: quantum chemical hybrid methods for calculating interaction energies, J. Phys. Chem. A 108 (2004) 107–112.
[67] E. Harder, A.D. MacKerell Jr., B. Roux, Electronic polarization effects and the membrane dipole potential, J. Am. Chem. Soc. 131 (2009) 2760–2761.
[68] R.M. Venable, A.J. Sodt, B. Rogaski, H. Rui, E. Hatcher, A.D. MacKerell Jr., R.W. Pastor, J.B. Klauda, CHARMM all-atom additive force field for sphingomyelin: elucidation of hydrogen bonding and of positive curvature, Biophys. J. 107 (1) (2014) 134–145.
[69] E.R. Hatcher, O. Guvench, A.D. MacKerell Jr., CHARMM additive all-atom force field for acyclic polyalcohols, acyclic carbohydrates, and inositol, J. Chem. Theory Comput. 5 (5) (2009) 1315–1327.
[70] E. Hatcher, O. Guvench, A.D. MacKerell Jr., CHARMM additive all-atom force field for aldopentofuranoses, methyl-aldopentofuranosides, and fructofuranose, J. Phys. Chem. B 113 (2009) 12466–12476.
[71] O. Guvench, E.R. Hatcher, R.M. Venable, R.W. Pastor, A.D. MacKerell Jr., CHARMM additive all-atom force field for glycosidic linkages between hexopyranoses, J. Chem. Theory Comput. 5 (2009) 2353–2370.
[72] E.P. Raman, O. Guvench, A.D. MacKerell Jr., CHARMM additive all-atom force field for glycosidic linkages in carbohydrates involving furanoses, J. Phys. Chem. B 114 (40) (2010) 12981–12994.
[73] S.S. Mallajosyula, K.M. Adams, J.J. Barchi, A.D. MacKerell, Conformational determinants of the activity of antiproliferative factor glycopeptide, J. Chem. Inf. Model. 53 (2013) 1127–1137.
[74] E. Hatcher, E. Sawen, G. Widmalm, A.D. MacKerell Jr., Conformational properties of methyl beta-maltoside and methyl alpha- and beta-cellobioside disaccharides, J. Phys. Chem. B 115 (2011) 597–608.

[75] D.S. Patel, R. Pendrill, S.S. Mallajosyula, G. Widmalm, A.D. MacKerell Jr., Conformational properties of α- or β-(1–6)-linked oligosaccharides: hamiltonian replica exchange MD simulations and NMR experiments, J. Phys. Chem. B 118 (2014) 2851–2871.
[76] M.C. Small, P. Lopes, R.B. Andrade, A.D. MacKerell Jr., PLoS Comput. Biol. 9 (2013) e1003113.
[77] S. Jo, T. Kim, V.G. Iyer, W. Im, CHARMM-GUI: a web-based graphical user interface for CHARMM, J. Comput. Chem. 29 (2008) 1859.
[78] S. Jo, K.C. Song, H. Desaire, A.D. MacKerell Jr., W. Im, Glycan reader: automated sugar identification and simulation preparation for carbohydrates and glycoproteins, J. Comput. Chem. 32 (2011) 3135–3141.
[79] B.R. Brooks, C.L. Brooks III, A.D. MacKerell Jr., L. Nilsson, R.J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A.R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R.W. Pastor, C.B. Post, J.Z. Pu, M. Schaefer, B. Tidor, R.M. Venable, H.L. Woodcock, X. Wu, W. Yang, D.M. York, M. Karplus, CHARMM: the biomolecular simulation program, J. Comput. Chem. 30 (2009) 1545–1614.
[80] J.C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R.D. Skeel, L. Kalé, K. Schulten, Scalable molecular dynamics with NAMD, J. Comput. Chem. 26 (2005) 1781–1802.
[81] K. Vanommeslaeghe, A.D. MacKerell Jr., Automation of the CHARMM General Force Field (CGenFF) I: bond perception and atom typing, J. Chem. Inf. Model. 52 (2012) 3144–3154.
[82] K. Vanommeslaeghe, E.P. Raman, A.D. MacKerell Jr., Automation of the CHARMM General Force Field (CGenFF) II: assignment of bonded parameters and partial atomic charges, J. Chem. Inf. Model. 52 (2012) 3155–3168.
[83] W. Yu, X. He, K. Vanommeslaeghe, A.D. MacKerell Jr., Extension of the CHARMM General Force Field to sulfonyl-containing compounds and its utility in biomolecular simulations, J. Comput. Chem. 33 (2012) 2451–2468.
[84] J. Chowdhary, E. Harder, P.E.M. Lopes, L. Huang, A.D. MacKerell Jr., B. Roux, A polarizable force field of dipalmitoylphosphatidylcholine based on the classical Drude model for molecular dynamics simulations of lipids, J. Phys. Chem. B 117 (2013) 9142–9160.
[85] J. Wang, P. Cieplak, J. Li, T. Hou, R. Luo, Y. Duan, Development of polarizable models for molecular mechanical calculations I: parameterization of atomic polarizability, J. Phys. Chem. B 115 (2011) 3091–3099.
[86] P.T. van Duijnen, M. Swart, Molecular and atomic polarizabilities: Thole's model revisited, J. Phys. Chem. A 102 (14) (1998) 2399–2407.
[87] E. Harder, V.M. Anisimov, T. Whitfield, A.D. MacKerell Jr., B. Roux, Understanding the dielectric properties of liquid amides from a polarizable force field, J. Phys. Chem. B 112 (2008) 3509–3521.
[88] M.L. Connolly, Analytical molecular surface calculation, J. Appl. Crystallogr. 16 (1983) 548–558.
[89] X. Zhu, A.D. Mackerell Jr., Polarizable empirical force field for sulfur-containing compounds based on the classical Drude oscillator model, J. Comput. Chem. 31 (2010) 2330–2341.
[90] K.J. Miller, Additivity methods in molecular polarizability, J. Am. Chem. Soc. 112 (1990) 8533–8542.
[91] P.E.M. Lopes, G. Lamoureux, B. Roux, A.D. MacKerell Jr., Polarizable empirical force field for aromatic compounds based on the classical Drude oscillator, J. Phys. Chem. B 111 (2007) 2873–2885.
[92] R. Bader, Atoms in Molecules: A Quantum Theory, Oxford University Press, USA, 1994.
[93] A. Savelyev, B. Roux, A.D. MacKerell Jr., Explicit inclusion of induced polarization in atomistic force fields based on the classical Drude oscillator model, in: Q. Cui, M. Meuwley (Eds.), Many-body Effects and Electrostatics in Biomolecules, Pan Stanford Publishing, 2014, (in press).
[94] J. Huang, B. Roux, A.D. MacKerell Jr., Recent advances in polarizable force fields for macromolecules: microsecond simulations of proteins using the classical Drude oscillator model, J. Phys. Chem. Lett. (2014) (in press).
[95] J. Huang, A.D. MacKerell Jr., Induction of peptide bond dipoles drives cooperative helix formation in the (AAQAA)3 peptide, Biophys. J. 107 (4) (2014) 991–997.
[96] R.B. Best, J. Mittal, M. Feig, A.D. MacKerell Jr., Inclusion of many-body effects in the additive CHARMM protein CMAP potential results in enhanced cooperativity of α-helix and β-hairpin formation, Biophys. J. 103 (5) (2012) 1045–1051.
[97] S. Lifson, A. Roig, On the theory of helix–coil transition in polypeptides, J. Chem. Phys. 34 (6) (1961) 1963–1974.
[98] I. Vorobyov, V.M. Anisimov, S. Greene, R.M. Venable, A. Moser, R.W. Pastor, A.D. MacKerell Jr., Additive and classical Drude polarizable force fields for linear and cyclic ethers, J. Chem. Theory Comput. 3 (3) (2007) 1120–1133.
[99] C.M. Baker, A.D. MacKerell Jr., Polarizability rescaling and atom-based Thole scaling in the CHARMM Drude polarizable force field for ethers, J. Mol. Model. 16 (3) (2010) 567–576.
[100] C.M. Baker, V.M. Anisimov, A.D. MacKerell Jr., Development of CHARMM polarizable force field for nucleic acid bases based on the classical Drude oscillator model, J. Phys. Chem. B 115 (3) (2011) 580–596.
[101] A. Savelyev, A.D. MacKerell Jr., Balancing the interactions of ions, water and DNA in the Drude polarizable force field, J. Phys. Chem. B 118 (2014) 6742–6757, http://dx.doi.org/10.1021/jp503469s.
[102] A. Savelyev, A.D. MacKerell Jr., All-atom polarizable force field for DNA based on the classical Drude oscillator model, J. Comput. Chem. 35 (16) (2014) 1219–1239.
[103] J.A. Lemkul, A. Savelyev, A.D. MacKerell Jr., Induced polarization influences the fundamental forces in DNA base flipping, J. Phys. Chem. Lett. 5 (2014) 2077–2083.
[104] X. He, P.E.M. Lopes, A.D. MacKerell Jr., Polarizable empirical force field for acyclic poly-alcohols based on the classical Drude oscillator, Biopolymers 99 (2013) 724–738, http://dx.doi.org/10.1002/bip.22286.
[105] D.S. Patel, X. He, A.D. MacKerell Jr., Polarizable empirical force field for hexopyranose monosaccharides based on the classical Drude oscillator, J. Phys. Chem. B (2014), http://dx.doi.org/10.1021/jp412696m.
[106] I.V. Vorobyov, V.M. Anisimov, A.D. MacKerell Jr., Polarizable empirical force field for alkanes based on the classical Drude oscillator model, J. Phys. Chem. B 109 (40) (2005) 18988–18999.
[107] P.E.M. Lopes, G. Lamoureux, A.D. MacKerell Jr., Polarizable empirical force field for nitrogen-containing heteroaromatic compounds based on the classical Drude oscillator, J. Comput. Chem. 30 (2009) 1821–1838.
[108] L. Huang, B. Roux, Automated force field parameterization for nonpolarizable and polarizable atomic models based on ab initio target data, J. Chem. Theory Comput. 9 (2013) 3543–3556.
[109] W. Jiang, D.J. Hardy, J.C. Phillips, A.D. MacKerell Jr., K. Schulten, B. Roux, High-performance scalable molecular dynamics simulations of a polarizable force field based on classical Drude oscillators in NAMD, J. Phys. Chem. Lett. 2 (2011) 87–92.
[110] C. Catlett, W.E. Allcock, P. Andrews, R. Aydt, R. Bair, N. Balac, B. Banister, T. Barker, M. Bartelt, P. Beckman, F. Berman, G. Bertoline, A. Blatecky, J. Boisseau, J. Bottum, J. Brunett, J. Bunn, M. Butler, D. Carver, J. Cobb, T. Cockerill, P.F. Couvares, M. Dahan, D. Diehl, T. Dunning, I. Foster, K. Gaither, D. Gannon, S. Goasguen, M. Grobe, D. Hart, D. Heinzel, C. Hempel, W. Huntoon, J. Insley, C. Jordan, I. Judson, A. Kamrath, N. Karonis, C. Kesselman, P. Kovatch, L. Lane, S. Lathrop, M. Levine, D. Lifka, L. Liming, M. Livny, R. Loft, D. Marcusiu, J. Marsteller, S. Martin, S. McCaulay, J. McGee, L. McGinnis, M. McRobbie, P. Messina, R. Moore, R. Moore, J.P. Navarro, J. Nichols, M.E. Papka, R. Pennington, G. Pike, J. Pool, R. Reddy, D. Reed, T. Rimovsky, E. Roberts, R. Roskies, S. Sanielevici, J.R. Scott, A. Shankar, M. Sheddon, M. Showerman, D. Simmel, A. Singer, D. Skow, S. Smallen, W. Smith, C. Song, R. Stevens, C. Stewart, R.B. Stock, N. Stone, J. Towns, T. Urban, M. Vildibill, E. Walker, V. Welch, N. Wilkins-Diehr, R. Williams, L. Winkler, L. Zhao, A. Zimmerman, TeraGrid: analysis of organization, system architecture, and middleware enabling new types of applications, in: L. Grandinetti (Ed.), High Performance Computing (HPC) and Grids in Action, vol. 16, IOS Press, Amsterdam, 2007.