



## Review article

## Molecular simulation of peptides coming of age: Accurate prediction of folding, dynamics and structures

Panagiota S. Georgoulia<sup>a,b</sup>, Nicholas M. Glykos<sup>c,\*</sup><sup>a</sup> Department of Chemistry and Biomedical Sciences, Faculty of Health and Life Sciences, Linnæus University, Kalmar, 39182, Sweden<sup>b</sup> Linnæus University Centre of Excellence “Biomaterials Chemistry”, Kalmar, 3912, Sweden<sup>c</sup> Department of Molecular Biology and Genetics, Democritus University of Thrace, University Campus, Alexandroupolis, 68100, Greece

## ARTICLE INFO

## Keywords:

Peptide simulations  
 Peptide folding  
 Peptide dynamics  
 Empirical force fields  
 Validation

## ABSTRACT

The application of molecular dynamics simulations to study the folding and dynamics of peptides has attracted a lot of interest in the last couple of decades. Following the successful prediction of the folding of several proteins using molecular simulation, foldable peptides emerged as a favourable system mainly due to their application in improving protein structure prediction methods and in drug design studies. However, their performance is inherently linked to the accuracy of the empirical force fields used in the simulations, whose optimisation and validation is of paramount importance. Here we review the most important findings in the field of molecular peptide simulations and highlight the significant advancements made over the last twenty years. Special reference is made on the simulation of disordered peptides and the remaining challenge to find a force field able to describe accurately their conformational landscape.

## 1. Overview

Molecular simulations have evolved to a powerful theoretical technique to study peptide structure and dynamics, as indicated by the plethora of references in the literature. The study of peptides thrived in the beginning of this millennium and still holds as an active field of research after 20 years. This interest is mainly due to the ability of foldable peptides to mimic essential characteristics of the much larger protein systems while at the same time maintaining the simplicity of smaller systems. This makes peptide systems not only easier to comprehend, but also cost efficient in terms of both computational power and experimental resources required. Their advantages as model systems have been eloquently presented in other reviews [24,63]. Herein, we try to underline their vital role in the development and advancement of force fields through numerous cycles of validation and optimisation that lead to what we perceive to be significant improvements in the ability of simulations to capture and reproduce the experimentally accessible physical reality.

## 2. Peptide folding simulations: two decades of continuous advancement

Historically, the study of peptides was prompted by the fact that they resemble the early events in protein folding and thus offer an

excellent opportunity to decipher protein-structure relationships [35,75,97]. Early studies suggested that peptides may serve as nucleation sites from where the folding is initiated [171,200]. This paved the way to a number of studies aiming to characterise the timescales of the early folding events and to model them computationally.

The speed limit of folding has been estimated to be  $N/100 \mu\text{s}$ , where  $N$  is the number of residues [96], with its actual value being dependant on a number of events such as the formation of hydrophobic core, hydrogen bonds, electrostatic interactions, solvation energy, conformational entropy, diffusion, etc. [126]. The timescales for the formation of basic structural elements have been placed to the sub-second time-scale [20,43]. The fastest early folding events were grouped by Gnanakaran et al. [63] based on the minimal sequence that can form each of these structures, from loop-closure events that can take place in the 10ns time-scale [103], to  $\alpha$ -helix formation which takes place in the 100–200ns time-scale [57,199] and to the slower  $\beta$ -hairpin formation that may require several microseconds [132]. The Eaton group contributed with extensive studies on the speed of folding, setting an upper limit for a 50-residue protein to  $1 \mu\text{s}$  [42,66]. This nano-to micro-second regime allowed a unique opportunity for a direct comparison between theoretical calculations and experimental measurements [27,95,103,202].

The combination of these relatively short folding timescales with the increase of the computational power available to the research

\* Corresponding author.

E-mail address: [glykos@mbg.duth.gr](mailto:glykos@mbg.duth.gr) (N.M. Glykos).<https://doi.org/10.1016/j.abbi.2019.01.033>

Received 14 September 2018; Received in revised form 23 January 2019; Accepted 28 January 2019

Available online 01 February 2019

0003-9861/ © 2019 Elsevier Inc. All rights reserved.

groups led to numerous peptide folding simulation studies which significantly advanced the field of peptide folding simulations. One of the pioneering studies was by Daura et al. that showed in atomistic detail the reversible folding of two  $\beta$ -peptides in solution [32]. Later, the Caflisch group showed the impressive, for that time, folding of a 20-residue peptide to a three-stranded antiparallel  $\beta$ -sheet through a cumulative simulation time of 4  $\mu$ s [49,50]. Likewise, Pande et al., illustrated the efficiency of using different starting conformations and high-temperature unfolding to study the folding pathway of a  $\beta$ -hairpin fragment of protein G [143]. Challenging short peptides with many charged residues (10 out of 21 amino acids) were successfully computationally folded to the native structure in an attempt to describe the folding pathway and estimate folding rates that were experimentally difficult to access [194].

In later years, the increase in available computational power made it possible for the molecular simulations to reach even higher time scales in the range of tens to hundreds of microseconds, approaching the folding times of longer peptides (20–60 amino acids) [125]. One of the first notable studies was the milestone 1  $\mu$ s single folding simulation of the 36 residue villin headpiece in explicit water by Duan and Kollman [38]. Shortly after that, the Pande group demonstrated the efficiency of using many short trajectories instead of a single long one to describe the folding pathway of small peptides like BBA and villin headpiece (23 and 36 residue long, respectively), an effort made possible through the development of distributed computing [45,104,167,175,205].

A few peptides served as preferable study systems due to their unique properties and abundance of experimental data, like the 20-residue peptide, Trp-cage, the smallest stably folded peptide showing two-state folding properties [28,151,170] and the WW domain (approximately 40 residues) of Pin1 protein [46,52]. The later one served especially as a workhorse for small all  $\beta$ -sheet peptides to examine force fields' bias towards helical conformations [53,131].

All-atom molecular simulations in explicit solvent have nowadays reached the impressive millisecond timescale thanks to the improvement of the parallelization algorithms [2], the highly optimised GPU implementations [55,159,180] and special-purpose supercomputers like Anton [163,164]. For an example of the latter, it was the increased simulation length afforded by the Anton machine, together with a better performing force field, that allowed the correct folding of both FiP35, the fastest folding variant of WW domains, and BPTI (58 residues) to experimental resolution [164], overcoming deficiencies of previous trials [52,53]. The Shaw group in 2011 took the whole field of molecular simulation a huge step forward by demonstrating the correct folding of 12 structurally diverse small proteins, ranging from 10 to 80 residues and from all structural classes [108]. Apart from the obvious significance of a successful *in silico* folding study, they demonstrated the transient formation of native-like structural elements already in the unfolded state, providing unprecedented insights into their folding mechanism.

The small size of the peptides is counteracted by their high flexibility and increased dynamics which made the probing of their full conformational potential extremely difficult with conventional unconstrained molecular dynamics simulations. Thus, and more often than not, enhanced methods are employed to study peptide systems. In fact most of the work cited in this review concerns studies that make use of some of the established accelerated dynamics techniques, such as the replica exchange molecular dynamics (REMD) [181], simulated tempering (ST) [120] and adaptive tempering [206] to name but few. The common concept in all of these methods is to enhance the conformational sampling through different biasing methods (for example umbrella sampling [182], metadynamics [100], and accelerated MD [68]) to achieve the desired long timescales and retrieve thermodynamic and kinetic properties for the system. Although an in-depth presentation of these techniques is beyond the scope of this review, we should emphasise the contribution of these methods to the field of peptide folding

[76,179,207] and refer the interested reader to some excellent reviews already available in the literature [36,122,130,141].

To summarise the progress outlined above, we believe that molecular simulation of peptides has reached the level of maturity where they can realistically describe in atomistic detail the structure and dynamics of peptides, justifying their naming as a computational microscope [37]. In the past, a folding simulation could have been characterised as successful based solely on its ability to locate the native (experimentally determined) structure. It is an indication of the maturity of the field that simply locating the native peptide structure through simulation is no longer sufficient for an application to be considered successful. The major question for recent applications is how realistically can the simulation describe the folding mechanism and the corresponding folding landscape, including the transiently stable conformations. Present-day molecular simulations are performing relatively well in characterising structurally stable peptide conformations and in estimating rates and free energies of folding [149]. However the unfolded state is still poorly understood. The inherent properties of the disordered state—with its multitude of conformations of similar energies which are separated by relatively low energy barriers—amplifies any small force field deficiencies but will most probably not prevent the native structure from being located. The ability to perform extremely long simulations of peptide systems will allow a more extensive sampling of the unfolded/disordered state and should, thus, allow further refinement of the empirical force fields.

### 3. Molecular simulations of health-related peptides

A great interest for the study of the structure and dynamics of peptides came from their applications in drug design, even leading to the creation of a new subfield aptly named “computational peptidology” [209]. Molecular dynamics are a powerful tool towards this means, whose contribution is increasingly being recognised [22,41]. Simulation-based methods are actively used both in early stages of drug discovery as well as in the final steps aiming to decipher binding mechanisms [208].

Peptides are physiologically involved in many biological processes with vital roles in metabolism and signalling. In particular, they are long-recognised to have a well-established role as anti-cancer, anti-microbial, anti-viral, anti-angiogenic and anti-inflammatory agents. Their inherent capability to modulate protein function increases dramatically their druggability potency [89]. On top of that, they offer enhanced activity and specificity which is translated in smaller dosages and lower toxicity. However, their susceptibility to proteases appears to be a major bottleneck, underlying the need for alternative administration routes to overcome their low bioavailability [44]. Significant improvements in the synthesis and purification techniques have made the production of peptides up to 100 amino acids a routine, opening new horizons for their large-scale health-related applications. A comprehensive presentation of the peptide-based drug market is exhaustively analysed in other reviews [110,188].

Significant advancements have also been made concerning the availability of computational tools aiming to assist peptide-based drug design studies [90,91]. However, most of these tools are limited in their general applicability by making the more often than not unjustified assumption that the peptides of interest will have stable (native-like) structures in water. In this connection, special mention must be made of the Rosetta suite of programs which have been catalytic in the field, with pioneering work in the *de novo* peptide (and protein) structure prediction accuracy as assessed by CASP (Critical Assessment of Structure Prediction) [21]. The Rosetta algorithm implemented the fundamental concept of peptide fragments, which are short fragments of known protein structures that are assembled by Monte Carlo methods to predict protein structures [7,172]. There are also many other *in silico* tools dedicated to structure prediction but only a handful are adapted for peptides. Their accuracy is limited and highly dependable on the

availability of experimental data. The most promising representative of this class is the Pepfold server that is a bioinformatics-based de novo approach to predict tertiary peptide structures from sequence [124]. The algorithm uses Hidden Markov Models and non-overlapping peptide fragments of four amino acids. It should be noted, however, that all of these methods and programs, and irrespectively of their detailed algorithmic basis, utilise empirical force fields to refine their proposed structures, thus highlighting the substantial interplay of peptides with the force fields even in protein structure prediction methods.

Health-related applications of peptides also involve the search for small foldable peptides with predefined structural characteristics. Identification of small stably folded and soluble peptides is not trivial [59]. Many small peptides (4–10 residues) of biological interest have been studied computationally but were found to be marginally stable in solution [171]. Well-known exceptions to this rule are the designed 10-residue peptides chignolin [77,98,160] and its variant, CLN025 [70,127,157] that adopt a unique and outstandingly stable  $\beta$ -hairpin structure in water.

The question that arises then, is how good are the empirical force fields in describing the structure and dynamics of peptides, especially of those that are not stably folded, or have more than one stable conformation. Are, for example, the force fields sensitive enough to predict the structural plasticity of the peptides, or, the sometimes pronounced effect of mutations on the peptide structure and dynamics? Force field improvement and validation has gone hand-in-hand with peptide simulations. In the following section we present a historical perspective of the evolution of the force fields highlighting the parallel paths that accurate peptide structure predictions and force field development followed through the years.

#### 4. Force fields: then and now

A number of force fields have been developed over the years for the simulations of biological macromolecules. These can be categorised in the families of CHARMM [114], AMBER [31], OPLS [85] and GROMOS [140]. The similarities and differences among them, especially regarding their parameterisation protocols can be found in other reviews [65,153]. In what follows, we review the evolution of the force fields from their first appearance to the present, highlighting the pivotal role of peptides in their improvement.

##### 4.1. CHARMM force field

Historically, the CHARMM (Chemistry at Harvard Macromolecular Mechanics) force field appeared in 1983 [23] and was officially released in the version CHARMM19 [134] (Fig. 1). The published parameters comprised the so called ‘extended’ atom types and van der Waals parameters, wherein hydrogen atoms were included as part of their attached heavy atoms for sulfur and carbon, whereas for nitrogen and oxygen the hydrogen atoms were treated explicitly, hence the name united-atom potentials. One of the major features was the application of the partial atomic charges to fit ab initio calculations, using the TIP3P water model [84] to calibrate the interactions. The potential function also included internal energy terms for bonds, angles, dihedrals, improper dihedrals as well as the non-bonded terms for van der Waals and electrostatics, using a truncation cutoff for the latter. This basic form of the energy function has been carried-over to almost all present-day empirical force fields.

CHARMM22 [113,114] was later introduced as an ‘additive’ protein force field that included explicit parameters for all atoms (rendering the term ‘extended atom’ obsolete), emphasising on the balance of interactions involving the protein and the solvent and in-between. This was mainly achieved through a refinement of the non-bonded interactions already implemented in the CHARMM19 version, by broadening the experimental and ab initio data used. Backbone parameters were based on the *N*-methylacetamide (NMA) and the alanine dipeptide

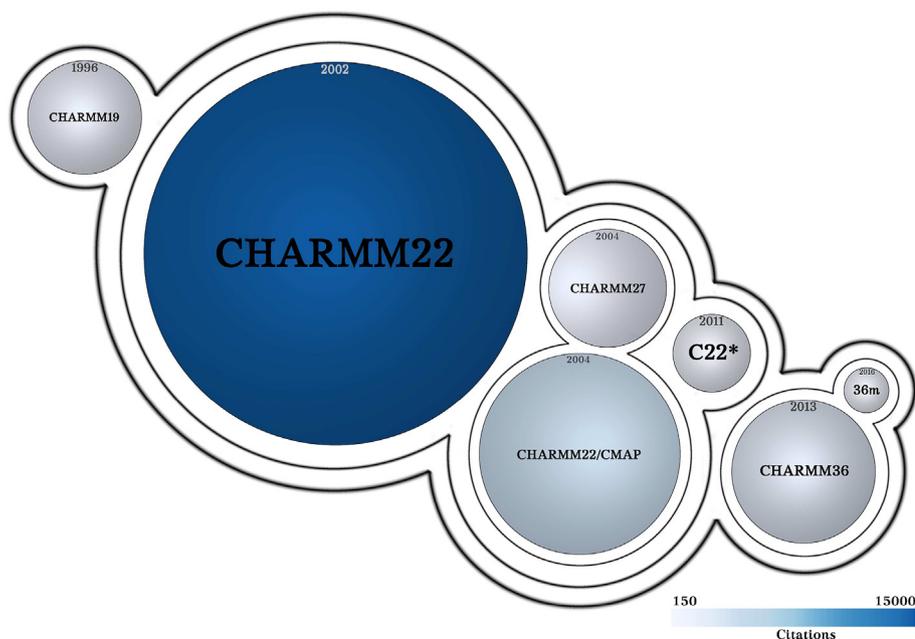
and the side chains were based on a number of small model compounds. Additive force fields do not allow however the change of the electrostatic parameters as a function of environment, neglecting the electronic polarisation phenomenon. The CHARMM22 protein parameters were maintained to the version that followed, CHARMM27, where only the nucleic acid and lipid parameters were updated together with parameters for a number of common ions [112].

This force field version remained prevalent for a decade until the introduction of the CHARMM22/CMAP version which incorporated the addition of a dihedral correction map (CMAP) [115,116]. CMAP is a two-dimensional grid of energy corrections in the dihedral space of  $\phi/\psi$  angles added to the potential energy equation, which improved the secondary structure propensities and compensated for the demonstrated helical bias of the CHARMM force field [178]. The parameters for the CMAP term were derived from a big collection of protein x-ray crystallographic data [116]. The CHARMM22/CMAP version had an improved performance [25], but deficiencies in the equilibrium between helical and extended conformations were noted which affected the ability of the force field to correctly fold small peptides, mostly due to an increased helical propensity [13,52,53]. Thus, further improvements were incorporated by correcting the backbone and side-chain dihedral angles to sample better the corresponding regions in the Ramachandran plot [19]. Corrections were based on experimental solution NMR data of weakly structured peptides for all residues except glycine and proline, for which QM-based corrections were made. We should note that this work was not based solely on alanine peptides but also on small helical and hairpin peptides.

In later years, only non-protein parameters regarding lipids, carbohydrates, drug-like and cofactor molecules were implemented, giving rise consequently to CHARMM27r, CHARMM35 and the CHARMM General Force Field (CGenFF). Together they resulted in the latest great release of CHARMM36 [19,81] and a few years later of CHARMM36m for intrinsically disordered proteins [82]. In parallel, version CHARMM C22\* [150] was presented that aimed towards a better helix-coil balance. This was achieved by replacing the CMAP correction for all residues except glycine and proline, and modifying the side-chain partial charges and torsional terms for residues aspartate, glutamate and arginine (following the philosophy of the ff99SB\* and ff03\* corrections in the AMBER family that is presented in the next section). A more comprehensive presentation on each component of the CHARMM all-atom additive force field together with the parameterisation philosophy and methodology and the recently released CHARMM Drude polarizable force field can be found elsewhere [111,187,210].

##### 4.2. AMBER force field

The AMBER (Assisted Model Building with Energy Refinement) force field first appeared also in the early 80s, sharing the united-atom idea [195] but was soon extended to an all-atom force field with recalibrated torsional and angle parameters based on experimental conformational energies [196] (Fig. 2). A decade of improvements in the parameters and the algorithm gave rise to a second generation force field, named ff94 [31]. Since then, the tradition has it that protein (and nucleic acid) force fields in the AMBER family are named with ‘ff’ followed by a two-digit number year. Up to this point, the parameterisation was focused on the gas phase behaviour, but the need to produce potentials that are suitable for condensed phase simulations became apparent for a more balanced modelling of biomolecules in solution. Unlike CHARMM though, the core parameterisation is different here, employing more adjustable empirical parameters that can be optimised. For example, the fixed atomic partial charges of CHARMM that are tightly connected to the TIP3P water model are not present here. Instead AMBER employed initially an ESP (electrostatic potential) fit for atomic centred charges and later an RESP fit (restrained ESP) where charges were fitted simultaneously to several conformations to achieve a better average behaviour. The dihedral

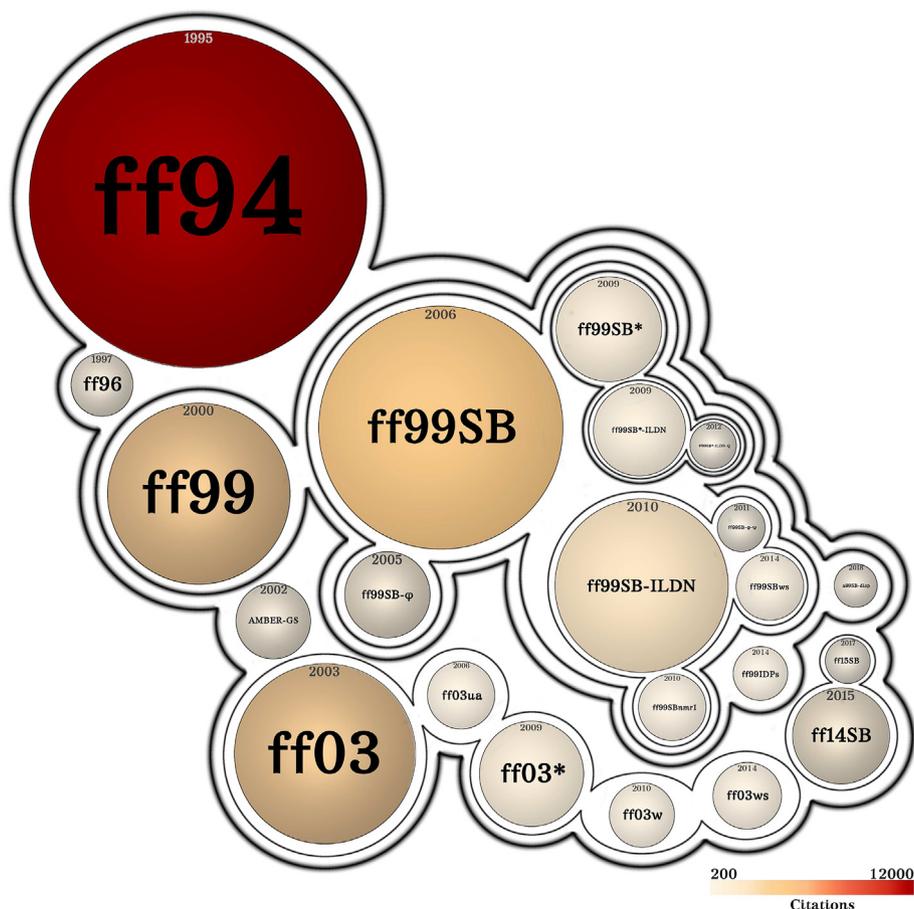


**Fig. 1.** CHARMM family of protein force fields. Circled tree map representation of the various protein-related force fields developed to present time. The size of the circle and the colour is analogous to the number of reported citations. Genealogically-related force fields are encircled in the same hierarchical level, which is represented by the thickness of the lines. Force-fields are ordered chronologically from left to right by year of appearance.

parameters were fit to relative QM energies of alternate rotamers of small molecules to represent several conformations of glycine and alanine. Unlike CHARMM, the AMBER force fields underwent many optimisations and refinements to reach a satisfactory level of accuracy, giving rise to a large number of variants.

The ff94 force field was shown to over-stabilise helical

conformations and over-estimate computed melting temperatures in peptide simulations, sometimes even bias helices over the native  $\beta$ -hairpin conformation [57,137,170]. The ff94 version was soon replaced by ff96 and C96 [93], which featured improved calculations for electrostatics and van der Waals interactions and better accounted of long-range effects. This resulted in better fitting to ab initio calculations on



**Fig. 2.** AMBER family of protein force fields. Circled tree map representation of the various protein force fields developed to present time as presented in Fig. 1.

small peptides and more balanced solvent-solvent and solute-solvent interactions. One of the main drawbacks of this version was a bias towards  $\beta$ -structures this time [74,88,138]. The next version ff99 [190] was yet another attempt to refit the backbone dihedral parameters using this time more representative structures of alanine and glycine in the form of tetrapeptides and dipeptides. At that time, generalised parameters for compounds beyond proteins and nucleic acids were also released. The new potential surfaces were significantly different of those of its predecessors.

Force field optimisation was assisted by the longer simulation times that became feasible, which only uncovered more force field deficiencies. This opened up to a decade of numerous almost in-parallel efforts to improve AMBER force fields, making it challenging to keep up with them and even more to choose the appropriate one for a particular study. Despite the extended optimisation of the ff99 version, AMBER force fields continued to suffer from an imbalance of secondary structural elements, over-stabilising  $\alpha$ -helical peptides, like their ff94 counterpart [57,137].

A few years later ff03 was released, a so called third-generation point-charge all-atom force field [39]. The major contribution was the fitting of the electrostatic potential and main-chain torsion parameters to QM calculations of small peptides in condensed phase with continuum solvent models. Initial testing showed a satisfactory balance between helical and extended conformations and a better description of the polyproline II region. A united-atom counterpart was also released, ff03ua [201], where aliphatic hydrogens of  $\alpha$ -carbon and aromatic hydrogens are explicitly represented whereas aliphatic hydrogens of side-chains are represented united to their carbon atom, to improve speed for computationally demanding cases like folding simulations. A close variant of ff03, ff03\* incorporated a correction to the backbone dihedral potential to fit experimental data of helix-coil transition, making it more transferable and able to fold peptides from both helical and beta structural classes [15]. The ff03w force field [16] is a slightly modified version of ff03\*, with a backbone dihedral potential correction that allows the usage of the more accurate TIP4P/2005 water model [1]. This updated force field offered a better description of the folding thermodynamics and of the unfolded state as tested in small peptides, whilst preserving its ability to fold to the native structure larger peptides like the Trp cage and the GB1 hairpin. It still suffered though from poor solvation, with less favourable solvation free energies and too favourable protein association. In an attempt to further improve the representation of these properties, the ff03ws force field [18] was released which essentially incorporated a tuning parameter for the Lennard-Jones protein-water interaction, specifically the water oxygen. The same scheme applied to the ff99SB force field, gave rise to version ff99SBws [18].

The ff99SB force field [79] was a parallel effort focused largely on improving the  $\phi/\psi$  dihedral terms of the previous ff94/ff99 energy functions, after showing that the inadequate backbone dihedral parameterisation was indeed responsible for the weaknesses of that generation of force fields. The major realisation here was that the two sets of dihedral backbone parameters introduced in ff94 version had to be both individually optimised for glycine and non-glycine residues. Glycine due to absence of a C $\beta$  atom needs only one set of dihedral parameters,  $\phi/\psi$  (defined as  $\phi = \text{C-N-C}\alpha\text{-C}$ ,  $\psi = \text{N-C}\alpha\text{-C-N}$ ), whereas all other amino acids that harbor a side-chain need also a  $\phi'/\psi'$  term (defined as  $\phi' = \text{C-N-C}\alpha\text{-C}\beta$ ,  $\psi' = \text{C}\beta\text{-C}\alpha\text{-C-N}$ ). Then the dihedral potential for the non-glycine residues will be the sum of them. This resulted in better agreement with the PDB data for the dihedral angles and consequently better description of the secondary structural preferences of small peptides as well as improved fitting to NMR observables of larger peptides, like trpzip2, trp cage, villin headpiece, ubiquitin and lysozyme.

In the same spirit, there were some efforts like the AMBER-GS version [57], where the  $\phi/\psi$  torsion potential was completely removed to fit better to experimental helix-coil parameters, but the  $\alpha$ -helical bias

was not removed as shown by long equilibrium ensemble simulations [177]. As a remedy, the ff99SB- $\phi'$  force field [177] was proposed where they combined the low helicity of the ff99 potential with the high helicity of the ff94, by removing the additional barriers on the  $\phi$  rotational degree of freedom. This version provided a better description of the helix-coil transition but its usage to non-helical peptides was not verified.

The fast —almost chaotic— production of even more versions of AMBER force fields continued unimpeded in the years to come. All versions from this period shared the core potential function of ff99SB which demonstrated superior performance to previous versions [13], proving the significance of the proper backbone dihedral parameterisation. Next in line, came ff99SB\* force field [15], which incorporated to 99SB potential the same backbone correction as the ff03\* version described above. This improved the helical bias and showed better agreement to NMR data of peptides but with poor thermodynamics description.

A different approach was introduced with the ff99SB-ILDN version [109], where the torsion potentials of side-chains were targeted this time. Especially four residue types, isoleucine (I), leucine (L), aspartate (D) and asparagine (N), were found to have considerably different rotamer distributions in the simulation in respect to the experimentally observed ones. So, the side-chain torsion potentials for these four residues were fitted to high-level QM calculations, improving considerably the agreement to NMR data for the rotamer distributions. Of course combinations of these optimisations surfaced, like ff99SB\*-ILDN [15] and later ff99SB\*-ILDN-Q [14], which combined the global backbone correction and the modified torsion parameters for residues ILDN, with refitted charges for residues D, E, K, R.

On another front, the Brüschweiler group introduced the concept of employing experimental NMR data that are commonly used to cross-validate a force field's performance, to optimise them instead. Thus they used a collection of NMR chemical shift data and RDCs for peptides to optimise and improve the ff99SB version using an energy-based reweighting to match the experimental data in an iterative manner, giving rise to the versions ff99SBnmr1 [105] and ff99SB- $\phi\psi$  [106] that showed superior performance when compared to its parent force field, ff99SB.

Around that time, it became clear that it was imperative to produce a catholic AMBER version with carefully selected corrections that can reproduce faithfully experimental data for peptides from all structural classes that the developers can responsibly recommend to general users of molecular simulations. Towards this direction, a preliminary version, ff12SB, and nowadays ff14SB [118] were officially released. Taking lessons from all the previous optimisation attempts, this version comprised a complete refit of the side-chain dihedral potential of all amino acids, including alternate protonation states for the ionizable ones, and empirical adjustments to the backbone dihedral potential, including the  $\phi$  rotational profile. The superiority of this updated version was tested for reproducibility of the secondary structural content and NMR parameters of peptides and proteins in solution. The ff12SB version was also optimised by adding a CMAP-like correction, ff12SB-cMAP, for simulations with implicit solvent models, when computational speed is a crucial factor [146].

Although the ff14SB is the one officially still recommended by the developers at the time of the writing, another version, ff15SB, came up for use in combination with the TIP3P-FB water model [192]. This version comprised a complete refitting of the bonded parameters of the parental ff99SB force field and was shown to retain the prediction accuracy of equilibrium properties while improving the accuracy of the temperature dependent ones [127]. Its broader application and establishment remains to be seen. For completeness, there are nowadays available force fields that can also describe small molecule ligands (General AMBER force field, GAFF), carbohydrates (GLYCAM), lipids (lipid14) but also amber-compatible parameters for post-translational modifications (Forcefield PTM) and ff15ipq version for implicit

polarised charges in explicit solvent [34].

#### 4.3. Force fields for disordered peptides

Much of the current focus in the community of force field development has turned towards the study of disordered peptides. An accurate description of the landscape of these structurally challenging peptides is of paramount importance due to the increasing acknowledgement of their abundance in protein structures as modular elements and their participation in many cellular processes [40]. Their main characteristic, the ability to interconvert between different conformations, also presents the greatest challenge which is the requirement to transform force fields that underwent decades of optimisation cycles to model well-folded peptides, to now perform comparably well even with disordered peptides.

Disordered peptides present the additional challenge of interpreting complex experimental observables that are averaged ensembles over transient conformations. The most common experimental techniques used for validation of the simulation data comprise NMR (nuclear magnetic resonance), usually amide proton exchange measurements, J-couplings, chemical shifts, NOEs (nuclear Overhauser effect) and RDC (residual dipolar coupling) constants, FRET (Förster resonance energy transfer) and SAXS (small angle X-ray scattering), that have been reviewed recently [18,26,72,80]. Infrared spectroscopy has emerged lately as a powerful tool to probe the conformational dynamics of the disordered states, providing experimental evidence on the secondary structures of sites that can be isolated by isotope labelling that can be compared with computationally modelled amide I spectra [48].

Naturally, force fields had to embrace the need to represent the disordered state as accurately as the well-structured state. The first attempt to produce such a force field was based on the application of the CMAP correction method to the dihedral potential of 8 disorder-promoting residues (A, G, P, R, Q, S, E, K), using as basis the ff99SB-ildn force field, giving rise to the version ff99IDPs [193] (Fig. 2). A later validation study using a set of 3 representative IDPs showed a quantitative agreement with the experimental NMR chemical shifts and more representative conformational sampling, though still suffering from over-stabilisation of structured conformations, most probably due to inadequate representation of polar interactions [203].

A more comprehensive effort was presented through the refined CHARMM36m (Fig. 1) force field [82] that tried to address a reported bias for overpopulation of the  $\alpha_L$  region of the Ramachandran plot [154]. The exhaustive benchmarking gave satisfactory fitting to experimental parameters for IDPs while retaining consistency with the CHARMM36 counterpart for folded peptides, folding kinetics and free energy calculations. It removed the oversampling of the  $\alpha_L$  region observed in CHARMM36, but also underestimated the  $\beta$  region as shown through simulations of peptides like chignolin and CLN025. The main issue addressed was the over-compactness, a common characteristic of all current generation empirical force fields when applied to studies of the disordered state.

In the same spirit, and inspired by previous optimisation efforts, a force field termed a99SB-disp was released (Fig. 2), the most recent to our knowledge empirical force field dedicated to accurately present both folded and unfolded conformational states [156]. This force field is based on the ff99SB-ILDN version with necessary modifications to employ the TIP4P-D water model. Its superior performance is attributed to (a) an iterative scheme to optimise the backbone and side-chain torsion parameters to represent better the protein-water van der Waals interactions inspired by the ff99SB\*-ILDN-Q version, and, (b) the modification of the strength of the Lennard-Jones term for carbonyl oxygen and amide hydrogen pairs. Its performance remains to be evaluated in the years to come.

Amyloid (A $\beta$ ) peptides and in particular their assembly into fibrils, have been one of the most challenging peptide systems to study with significant difficulties encountered with both the experimental and

computational approaches. The interested reader is referred to some excellent reviews of these Alzheimer's disease-related peptides [133,183]. A large number of studies appeared which attempted to computationally characterise the peptides' equilibrium conformations using the most advanced force fields. The great discrepancy among the predicted structures of the monomers as well as of the various oligomers only highlights the difficulty to describe the conformational preferences of disordered peptides and the force field bias that needs yet to be addressed [136]. The differences between the various force fields, sometimes in the context of different water models as well, have been attributed to differing secondary structure biases as well as the intrapeptide hydrogen bonding, with helix-overestimating force fields, like AMBER99SB and CHARMM22-CMAP, having the poorest agreement to experimentally observed populations [174]. Later studies with improved force field versions accompanied by different water models, such as OPLS-AA/TIP3P, AMBER99SB-ILDN/TIP4P-Ew and CHARMM22\*/TIP3P demonstrated strongly converged structural properties for these type of disordered peptides [158] and improved agreement with experimental J-coupling constants, chemical shifts and CD data [176]. These encouraging findings indicate that force field improvements may be heading in the correct direction. However, the force field validation is difficult due to the limited availability of experimental observables for disordered peptide systems. This is particularly concerning when force field performance is found to be dependent on the peptide system per se. For instance, the CHARMM36/TIP3P combination was the most accurate in the case of A $\beta$ 10-40 peptide but not in the case of A $\beta$ 1-40 and A $\beta$ 1-42 peptides [173]. In a different study, the most recent versions AMBER14SB and CHARMM22\* with TIP3P water model showed a helical overestimation compared to CD data, whereas OPLS-AA and AMBER99SB-ILDN with the same water model represented conformational ensembles closer to experiment [119]. Clearly further studies are needed to determine the source of the force field discrepancies, but more importantly to also comprehend whether the limited accuracy for disordered peptide systems is due to the interplay between the force field and water model parameters.

#### 4.4. Other biomolecular force fields

The force field families of CHARMM and AMBER are the ones that underwent through heavy optimisation and validation by the developers but also the rest of the community in the peptide folding and dynamics field. For completeness we should also mention two additional and significant contributions, those of the OPLS and GROMOS force fields.

The OPLS (optimised potentials for liquid simulations) made its appearance in the 80s as well (OPLS-ua) and was initially intended for simulation of liquid-state properties of water and other organic liquids [84] (Fig. 3). A distinct feature was the emphasis given to model the non-bonded interactions to fit liquid-state thermodynamic properties, like density and heat of vaporisation. The first release was a combined AMBER/OPLS force field [86] (AMBER ff94 was largely inspired by OPLS) and was later updated to the all-atom version OPLS-AA [85]. An update of it came out a few years later, where side-chain potentials were reparameterized based on QM-data of small peptides [87]. The major contribution from this effort was, and still is, the development of the water models TIP3P and TIP4P that are heavily used to the present. It was only very recently that new OPLS versions emerged, like OPLS2.1 and OPLS3, that are mostly optimised towards small molecules and protein-ligand binding studies [69].

Closing this section, we should refer to GROMOS (Groningen molecular simulation) that appeared in the 90s together with the homonymous computer simulation program with the most cited version being GROMOS96 [186] (Fig. 4). The newest release appeared a few years ago as GROMOS11 with the parameter sets 54A7 [161] and now 54A8 [155]. Its inferior performance compared to other force fields in

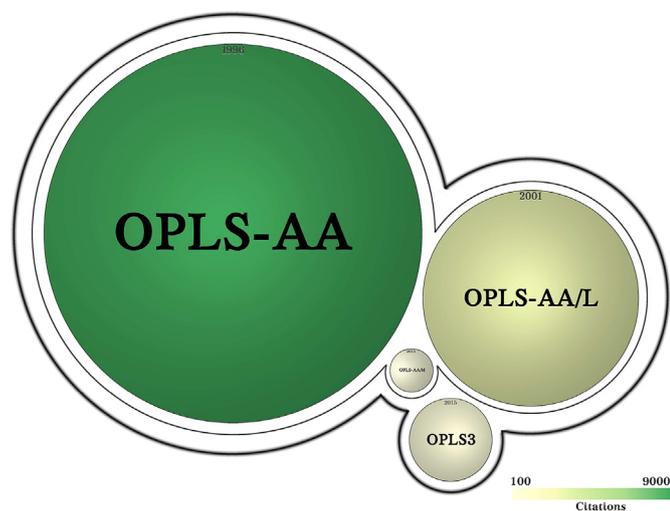


Fig. 3. OPLS family of protein force fields. Circled tree map representation of the various protein force fields developed to present time as presented in Fig. 1.

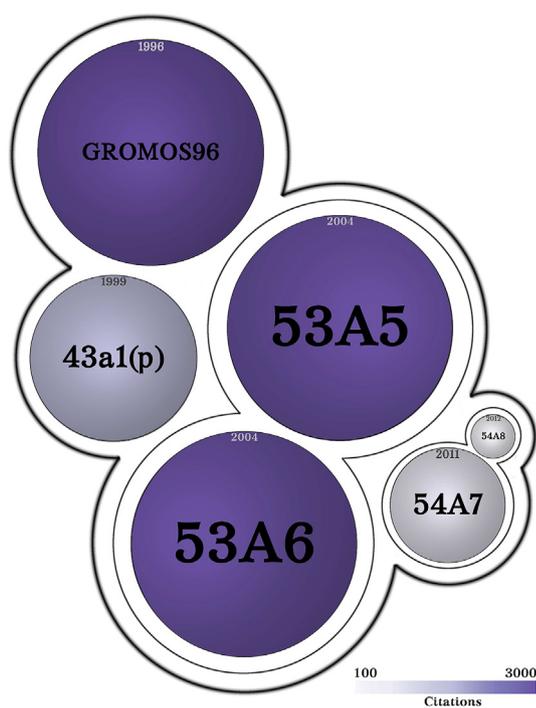


Fig. 4. GROMOS family of protein force fields. Circled tree map representation of the various protein force fields developed to present time as presented in Fig. 1.

comparative validation studies of peptide folding and dynamics prevented its broader use in the specified field of peptide simulations and will not be further presented here.

#### 4.5. Water models for explicit representation of the solute in peptide folding simulations

This review is focused on peptide folding simulations that comprise explicit representation of the solvent. Therefore, the ability of the force fields to describe accurately the conformational landscape of the peptides is inextricably dependent on the ability of the associated water model to describe the physical properties of water molecules in the liquid phase. There are numerous water potentials that have been developed so far, but we shall only mention here the empirical potentials

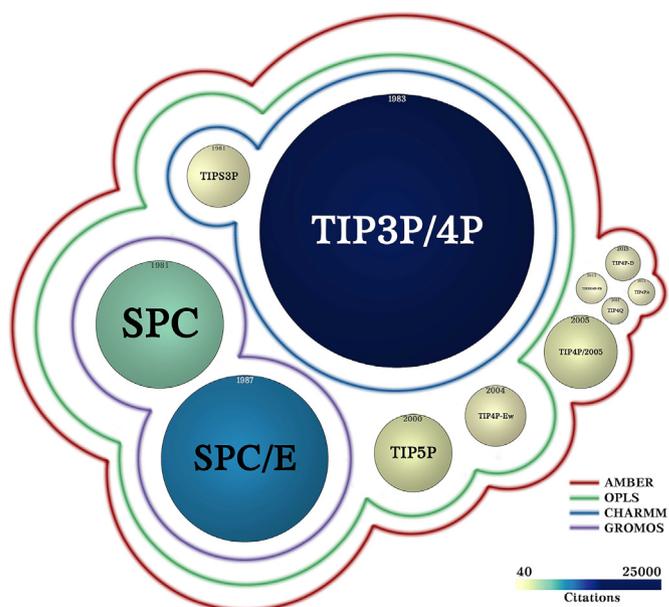


Fig. 5. Water models for biomolecular simulations. Circled tree map representation of the water potentials used in simulations of proteins and peptides. The size and colour of the circles follows their popularity as expressed by the number of citations. Encircled are the models that can be used in conjunction with a specific family of force fields, with consecutive levels of hierarchy represented by the line colour.

that are used in biomolecular simulations (Fig. 5). Detailed presentation of all available water models and their comparative performance can be found elsewhere [30,64,73,139,142,185,189].

Their first appearance coincided with the early protein force field versions in the 80s and they all share the idea of a rigid water monomer with the non-bonded interactions described by 3, 4, or 5 sites composed of Coulombic terms for the intermolecular interaction pairs and a Lennard-Jones term for oxygen atoms upon dimerization. Historically, two quite similar 3-site models were introduced in parallel in 1981. These were the TIP3 (transferable intermolecular potential) [83] model, and the SPC (simple point charge) [11] model, two of the most popular potentials for water molecules in molecular simulations even to date [153]. The main difference between TIP and SPC models is the tetrahedral shape geometry adapted by the latter and its ability to reproduce the experimental radial distribution function (including the second peak) and the self-diffusion coefficient. Later the SPC/E [10] model was released that comprised an additional polarisation correction leading to better representation of the density, diffusion coefficient and dielectric constant for simulations of liquid water. Together these features made the SPC models more appropriate for modelling the bulk properties of water [121].

Later reparameterisation of TIP3 and addition of a Lennard-Jones term to the hydrogen atoms as well, lead to improved representation of the density of liquid water and gave rise to what became the standard model for protein simulations, the TIP3P [84] model. Concurrently with TIP3P, its 4-site version was released, the TIP4P model, which also incorporated terms for the bisector of the H–O–H angle allowing a better electrostatic distribution. The TIP4P model provided a better description of most of the water's properties (like the phase diagram) but not of its dielectric constant. The increased computational cost (imposed by the larger number of interactions to be calculated) prohibited its wider use in the early peptide folding studies. The most crucial factor however for the wider application of TIP3P was its coupling to the vastly popular CHARMM22 force field. The 5-site interaction version, TIP5P [117] appeared in 2000, which replaced the single negative charge on the bisector angle of TIP4P with two single negative

charges on the lone-pair electrons. The truncated tetrahedral geometry showed excellent representation of the structure of water, matching the density and radial distribution properties, but suffered in describing the gas phase. The increased complexity also significantly increased the computational requirements. As a result, SPC and TIP4P models provided a satisfying description of liquid's water structural and thermodynamic parameters at a much lower computational cost.

The scenery changed with the increase in the available computational power but most importantly the appearance of AMBER force field versions that gave more freedom on the choice of water model (RESP-fitted atomic charges). All aforementioned water models utilise truncated Coulombic interactions. An improved treatment of the long-range interactions with particle-mesh Ewald summation gave rise to the TIP4P-Ew [78] model which improved the representation of liquid water's thermodynamic parameters over a large temperature range but without adding significantly to the computational cost. On the same spirit was based the parameterisation of the TIP4P/2005 [1] model which reproduces temperature-dependent properties but underestimates the dielectric constant. The TIP4P/ε [56] model is fine tuned to match exactly that property in a large temperature regime. An equivalent version is the TIP4Q [4] model that includes an additional charge to increase the dipole moment, but at an additional computational expense compared to other TIP4P-based models. The common characteristic of these water models is the larger enthalpy of solvation that leads to stronger interactions between the water and the embedded protein. It is for that reason that they have been lately employed in folding simulations of disordered less hydrophobic peptides, facilitating a more accurate description of the expanded unfolded state. On the same front, the TIP3P-FB and TIP4P-FB [191] models were introduced by applying the so called ForceBalance method that combines reference data produced both experimentally and theoretically to derive parameters for force fields. The resultant models, and in particular the TIP4P-FB, manage to reproduce most of the kinetic and thermodynamic properties of water, like the viscosity, diffusion coefficient, dielectric constant, radial distribution function and heat of vaporisation, in a variety of temperatures and pressures.

The take-home message here is that there is no perfect water model that encapsulates all of the experimental physical properties of water. Furthermore, the general users are highly restricted in their choice by the compatibility to the protein force field of choice and whether it is implemented in a molecular mechanics program. For example, the molecular dynamics simulation engine NAMD [147] only supports TIP3P- and TIP4P-based models from the whole panel of non polarizable water models. Gromacs [12,184] on the other hand supports SPC, SPC/E, TIP3P, TIP4P and TIP5P. It has thus become common knowledge in the field that the GROMOS force fields should better be paired with the SPC water models, OPLS force fields are recommended to be used with TIP4P, whilst the CHARMM family of force fields should be paired with the TIP3P model to avoid inconsistencies. AMBER force fields, especially the latest versions, are more versatile. Despite the relatively poor performance of some water models with respect to the description of the bulk solvent properties, their success in the folding of proteins and peptides appears to indicate that folding per se may not be as sensitive to the properties of bulk solvent. This finding is probably the result of the protein force fields being developed given the available water models, so their deficiencies could be masked in the empirical parameterisation of the force fields. As the observation that the main deficiency of current generation of force fields is the poor description of the protein-water interactions, an effort is underway to develop force fields that are parameterized in combination with the newly reported water models. In that spirit the TIP4P-D [148] was recently released in an effort to recapitulate the more extended conformational ensembles of the disordered state by increasing the strength of the dispersion interaction (LJ-term) (see also section 6. The future). As is always the case with molecular simulation, it will be the extensive testing by the community that will define which of these models will survive the acid

test of a quantitative comparison with the experimental data, including the ability to accurately describe the kinetics and thermodynamics of peptide folding.

## 5. The unceasing quest for a balanced biomolecular force field

The previous sections clearly demonstrated that one of the main contributions of peptide simulations is their continuing usage as test systems for the optimisation, correction and validation of empirical force fields [5,61,107,123,175]. One of the recurring challenges through all these studies has been to convincingly quantify force field accuracy, especially in connection with the fact that the relatively limited simulation timescales attainable are insufficient for guaranteeing a faithful configurational sampling of the peptides' folding landscapes [51]. Soon it became clear in the literature that the initial force field versions from all families were suffering from helical bias [13,51,53,62,79,123,137,165,178,204]. A common remedy to tackle the helical bias of the early force fields was to choose one with a bias towards the major secondary structure of the protein under study. This was obviously highly unsatisfactory and proved indeed ineffective, as demonstrated by the folding simulations of villin and Pin1 WW domain mentioned earlier with the CHARMM22/CMAP force field [115,116]. In these studies, villin folded to its native structure [54] but the WW domain, whose native structure is a 3-stranded β-sheet, did not [52]. The reason, of course, was that as far the force field was concerned a misfolded conformation had lower free energy than the native structure [47,53]. Studies such as these highlight how small force field inaccuracies can have an additive effect and can result in stabilising misfolded non-native structures.

AMBER's ff99SB was the first successful version to convincingly compensate for that and displayed superior performance over the rest of the force fields of its time: literature consistently provided overwhelming evidence of significantly improved balance of secondary structural elements and reasonable agreement with experimental data through multiple comparative validation studies comprising from short glycine and alanine peptides to longer 10–20 residue peptides (disordered, helical, beta) up to small proteins like lysozyme and ubiquitin [8,58,79,92,94,99,102,123,144,145,168,168,169,197,198]. Follow-up studies suggested two more force fields, ff99SB-ILDN-φ and ff99SB-ILDN-NMR, with significantly improved agreement with NMR observables, highlighting though that large uncertainties accompany this agreement: statistical calculated errors in the validation were in the margin of systematic experimental errors of these parameters [9].

The newer versions that surfaced, ff03/ff03\* and ff99SB\*-ILDN from AMBER, and CHARMM22/CMAP and CHARMM22\* from CHARMM, all succeeded in predicting the native states and folding rates (for example the case of the model protein villin headpiece), but displayed distinct discrepancies in the folding mechanism and in the description of the unfolded state in particular, with ff99SB\*-ILDN and CHARMM22\* being closer to the experiments [150]. CHARMM22\* showed superior performance over its counterparts in the CHARMM family in describing the free energy landscape of the β-hairpin GB1 [71]. Similarly, ff99SB and its variants (ff99SB\*, ff99SB-ILDN, ff99SB\*-ILDN), ff03, ff03\* and the GROMOS96 (43a1p,53a6) succeeded in finding the residual β-hairpin preference of the disordered 16-mer Nrf2-derived peptide but with extremely spurious results that were temperature dependent [29]. In accordance, a parallel study with ff99SB-ILDN, ff99SB\*-ILDN, CHARMM22/CMAP and CHARMM22\* provided accurate description of the native state of ubiquitin, GB3, villin and WW domain, but force fields with better helix-coil balance, ff99SB\*-ILDN, ff03\* and CHARMM22\* had superior performance in the peptide systems [107]. Our personal recent work also lends further support that for structurally challenging peptide systems, the ff99SB\*-ILDN is outperforming its counterparts in the AMBER 99SB family [3,60,162].

To summarise, our take on the current literature is that from the set of all extensively tested and validated force fields, there are nowadays

versions from both AMBER and CHARMM families that are shown to be balanced and well-performing in a wide variety of tested systems, including cases of intrinsically disordered peptide systems, like the RS repeats [154]. All that being said, we present in the next section our views concerning the remaining challenges towards a universally useful biomolecular force field.

## 6. The future

We believe that a bird's eye view of the recent literature leaves little doubt: peptide simulations followed the rest of the molecular dynamics field in making a tremendous progress towards accurate and physically relevant simulations. The development of new algorithms, the production and validation of new robust and transferable force fields, together with the increase of computational power available to the research groups, completely transformed the field. Indeed, it is no longer an exaggeration to state that we have reached the point where molecular simulations can faithfully predict and reproduce the peptides' native states, folding rates and in exceptional cases even provide atomistic description of the folding mechanism in unprecedented detail, often unreachable by experiments. Having said that, there are still a significant number of unresolved issues and deficiencies, some of which are discussed in the paragraphs that follow.

Peptides with weak structural propensities are difficult systems to track computationally. The major challenge here is to describe, without overestimating, the residual (if any) secondary structure present in the unfolded state, whilst retaining the folding performance for the folders. Some recently reported issues concern the observed temperature dependence of the conformational preferences of peptides and the weak folding cooperativity, whose combination results in a relatively poor description of the kinetics and thermodynamics [33]. As discussed in the previous section on water potentials, the source of this “overcompactness” compared to the experimental observables (SAXS, FRET, Rg) appears to be the inaccurate description of the solvation of the unfolded chain, possibly due to systematically underestimating the dispersion parameter in the protein-water interactions [18,72,148]. A number of approaches have been proposed to remedy this, for example applying a scaling factor to the van der Waals interaction parameter [18], or re-parameterizing the Lennard-Jones parameter for water's oxygen (TIP4P-D water model [148]) or hydrogen (mTIP3P) [82]. These approaches improved the agreement with the experimental observables for some peptide cases but not for others, rendering this approach not universally applicable. For example, recent studies showed that the more polar TIP3P model favours more compact structures [176]. On the other hand, the treatment of the protein-water interactions by the TIP4P-Ew and TIP4P/2005 water models, eases the hydrophobically-driven collapse and improves sampling of the disordered state for peptides like Trp-cage and GB1 [17], with TIP4P/2005 giving less collapsed conformations [135]. The latest of the released water models, TIP4P-D matched better the experimental Rg in a set of disordered peptides [148].

It has, thus, been realized that a new direction for further improvement of the force fields is the one aiming towards an accurate description of the solvation effect and better accounting for the long-range interactions. It is probably not surprising that the newest of the empirical protein force field versions are adapting to new water models that provide a better description of the total physical properties of the solvent. However, obtaining a balance between the water model and the protein force field is not a trivial exercise, especially with respect to peptide folding and dynamics. The bulk properties of the solvent are experimentally tractable and thus it is straightforward to validate the available water models on how well they describe the solvent's molecular properties. However, the impact of the interactions between the solvent and the biomolecules on the conformation of the latter is still poorly understood and needs further exploration. A much different approach has been introduced with a force field based on the Kirkwood-

Buff theory (KBFF) [152] to retrieve force field parameters that reproduce the microscopic characteristics of the solution to better describe the protein-water interactions. This force field was found to compensate for the overcompactness and match the experimentally found dimensions of disordered peptides [128], still though at the expense of partially unfolding peptides like GB1 [129]. However, this was partially rescued by increasing the interaction cut-off to strengthen the contribution of the long range interactions, like in the case of AMBER99SB\*-ILDN with the TIP4P-D water model for Nup peptides [148].

Polarizable force fields might be an alternative route by accounting for the environmental effects. Their main asset is the ability to describe changes of the charge distribution upon conformational changes of the molecules. This can be achieved via fluctuating charge models, Drude oscillator-based models [101], inducible dipole models or multipole electrostatics, like in AMOEBA [166], all of which are detailed in other reviews [6,67]. Their general applicability is hindered by the large computational burden of the additional electrostatic interatomic forces calculations, barely reaching the 100ns timescale even for small peptide systems.

The difficulty in approaching all those unresolved matters in a meaningful way is that on one hand the disordered state is heavily undersampled computationally and on the other hand complex experimental data of disordered peptides can have ambiguous interpretations. We believe that peptides have and will continue to serve as an acid test for the performance of molecular simulations and force fields to reproduce the physical reality: small deficiencies in the force field parameterisation can be shielded in the larger well-folded protein context. Further improvements could be of substantial importance towards more accurate modelling of disease-related issues, like mutation effects, misfolding, aggregation or alternative enzyme conformations that affect drug binding. We hope that this review of the literature regarding molecular simulations of peptides and associated force fields will aid the readers to wisely select a force field suitable for a certain application in order to achieve high quality results.

## Notes

The authors declare no competing financial interest.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.abb.2019.01.033>.

## References

- [1] J.L. Abascal, C. Vega, A general purpose model for the condensed phases of water: tip4p/2005, *J. Chem. Phys.* 123 (2005) 23450523.
- [2] M.J. Abraham, T. Murtola, R. Schulz, S. Páll, J.C. Smith, B. Hess, E. Lindahl, Gromacs, High performance molecular simulations through multi-level parallelism from laptops to supercomputers, *SoftwareX* 1 (2015) 19–25.
- [3] T. Adamidou, K.-O. Arvaniti, N.M. Glykos, Folding simulations of a nuclear receptor box-containing peptide demonstrate the structural persistence of the IxxII motif even in the absence of its cognate receptor, *J. Phys. Chem. B* 122 (1) (2017) 106–116.
- [4] J. Alejandro, G.A. Chapela, H. Saint-Martin, N. Mendoza, A non-polarizable model of water that yields the dielectric constant and the density anomalies of the liquid: tip4q, *Phys. Chem. Chem. Phys.* 13 (44) (2011) 19728–19740.
- [5] A.E. Aliev, D. Courtier-Murias, Experimental verification of force fields for molecular dynamics simulations using gly-pro-gly-gly, *J. Phys. Chem. B* 114 (38) (2010) 12358–12375.
- [6] C.M. Baker, Polarizable force fields for molecular dynamics simulations of biomolecules, *Wiley Interdiscipl. Rev.: Comput. Mol. Sci.* 5 (2) (2015) 241–254.
- [7] D. Baker, A. Sali, Protein structure prediction and structural genomics, *Science* 294 (5540) (2001) 93–96.
- [8] A.S. Baltzis, N.M. Glykos, Characterizing a partially ordered miniprotein through folding molecular dynamics simulations: comparison with the experimental data, *Protein Sci.* 25 (3) (2016) 587–596.
- [9] K.A. Beauchamp, Y.-S. Lin, R. Das, V.S. Pande, Are protein force fields getting better? a systematic benchmark on 524 diverse nmr measurements, *J. Chem. Theor. Comput.* 8 (4) (2012) 1409–1414.

- [10] H. Berendsen, J. Grigera, T. Straatsma, The missing term in effective pair potentials, *J. Phys. Chem.* 91 (24) (1987) 6269–6271.
- [11] H.J. Berendsen, J.P. Postma, W.F. van Gunsteren, J. Hermans, Interaction models for water in relation to protein hydration, *Intermolecular Forces*, Springer, 1981, pp. 331–342.
- [12] H.J. Berendsen, D. van der Spoel, van Drunen, R. Gromacs, A message-passing parallel molecular dynamics implementation, *Comput. Phys. Commun.* 91 (1995) 1–3 43–56.
- [13] R.B. Best, N.-V. Buchete, G. Hummer, Are current molecular dynamics force fields too helical? *Biophys. J.* 95 (1) (2008) L07–L09.
- [14] R.B. Best, D. de Sancho, J. Mittal, Residue-specific  $\alpha$ -helix propensities from molecular simulation, *Biophys. J.* 102 (6) (2012) 1462–1467.
- [15] R.B. Best, G. Hummer, Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides, *J. Phys. Chem. B* 113 (26) (2009) 9004–9015.
- [16] R.B. Best, J. Mittal, Protein simulations with an optimized water model: cooperative helix formation and temperature-induced unfolded state collapse, *J. Phys. Chem. B* 114 (46) (2010) 14916–14923.
- [17] R.B. Best, J. Mittal, Free-energy landscape of the gb1 hairpin in all-atom explicit solvent simulations with different force fields: similarities and differences, *Proteins: Struct. Func. Bioinf.* 79 (4) (2011) 1318–1328.
- [18] R.B. Best, W. Zheng, J. Mittal, Balanced protein–water interactions improve properties of disordered proteins and non-specific protein association, *J. Chem. Theor. Comput.* 10 (11) (2014) 5113–5124.
- [19] R.B. Best, X. Zhu, J. Shim, P.E. 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  $\phi$ ,  $\psi$  and side-chain  $\chi_1$  and  $\chi_2$  dihedral angles, *J. Chem. Theor. Comput.* 8 (9) (2012) 3257–3273.
- [20] O. Bieri, J. Wirz, B. Hellrung, M. Schutkowski, M. Drewello, T. Kiefhaber, The speed limit for protein folding measured by triplet–triplet energy transfer, *Proc. Natl. Acad. Sci. Unit. States Am.* 96 (17) (1999) 9597–9601.
- [21] R. Bonneau, D. Baker, Ab initio protein structure prediction: progress and prospects, *Annu. Rev. Biophys. Biomol. Struct.* 30 (1) (2001) 173–189.
- [22] D.W. Borhani, D.E. Shaw, The future of molecular dynamics simulations in drug discovery, *J. Comput. Aided Mol. Des.* 26 (1) (2012) 15–26.
- [23] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. a. Swaminathan, M. Charmm Karplus, A program for macromolecular energy, minimization, and dynamics calculations, *J. Comput. Chem.* 4 (2) (1983) 187–217.
- [24] C.L. Brooks, Protein and peptide folding explored with molecular simulations, *Accounts Chem. Res.* 35 (6) (2002) 447–454.
- [25] M. Buck, S. Bouguet-Bonnet, R.W. Pastor, A.D. MacKerell Jr., Importance of the cmap correction to the charmm22 protein force field: dynamics of hen lysozyme, *Biophys. J.* 90 (4) (2006) L36–L38.
- [26] V. Burger, T. Gurry, C. Stultz, Intrinsically disordered proteins: where computation meets experiment, *Polymers* 6 (10) (2014) 2684–2719.
- [27] T. Cellmer, M. Buscaglia, E.R. Henry, J. Hofrichter, W.A. Eaton, Making connections between ultrafast protein folding kinetics and molecular dynamics simulations, *Proc. Natl. Acad. Sci. Unit. States Am.* 108 (15) (2011) 6103–6108.
- [28] S. Chowdhury, M.C. Lee, G. Xiong, Y. Duan, Ab initio folding simulation of the trp-cage mini-protein approaches nmr resolution, *J. Mol. Biol.* 327 (3) (2003) 711–717.
- [29] E.A. Cino, W.-Y. Choy, M. Karttunen, Comparison of secondary structure formation using 10 different force fields in microsecond molecular dynamics simulations, *J. Chem. Theor. Comput.* 8 (8) (2012) 2725–2740.
- [30] G.A. Cisneros, K.T. Wikfeldt, L. Ojamae, J. Lu, Y. Xu, H. Torabifard, A.P. Bartok, G. Csanyi, V. Molinero, F. Paesani, Modeling molecular interactions in water: from pairwise to many-body potential energy functions, *Chem. Rev.* 116 (13) (2016) 7501–7528.
- [31] W.D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Merz, 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.
- [32] X. Daura, K. Gademann, B. Jaun, D. Seebach, W.F. Van Gunsteren, A.E. Mark, Peptide folding: when simulation meets experiment, *Angew. Chem. Int. Ed.* 38 (1999) 1–2 236–240.
- [33] R. Day, D. Paschek, A.E. Garcia, Microsecond simulations of the folding/unfolding thermodynamics of the trp-cage miniprotein, *Proteins: Struct. Func. Bioinf.* 78 (8) (2010) 1889–1899.
- [34] K.T. Debiec, D.S. Cerutti, L.R. Baker, A.M. Gronenborn, D.A. Case, L.T. Chong, Further along the road less traveled: amber ff15ipq, an original protein force field built on a self-consistent physical model, *J. Chem. Theor. Comput.* 12 (8) (2016) 3926–3947.
- [35] S. Doniach, P. Eastman, Protein dynamics simulations from nanoseconds to microseconds, *Curr. Opin. Struct. Biol.* 9 (2) (1999) 157–163.
- [36] U. Doshi, D. Hamelberg, Towards fast, rigorous and efficient conformational sampling of biomolecules: advances in accelerated molecular dynamics, *Biochim. Biophys. Acta Gen. Subj.* 5 (2015) (1850) 878–888.
- [37] R.O. Dror, R.M. Dirks, J. Grossman, H. Xu, D.E. Shaw, Biomolecular simulation: a computational microscope for molecular biology, *Annu. Rev. Biophys.* 41 (2012) 429–452.
- [38] Y. Duan, P.A. Kollman, Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution, *Science* 282 (5389) (1998) 740–744.
- [39] Y. Duan, C. Wu, S. Chowdhury, M.C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, et al., A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations, *J. Comput. Chem.* 24 (16) (2003) 1999–2012.
- [40] A.K. Dunker, J.D. Lawson, C.J. Brown, R.M. Williams, P. Romero, J.S. Oh, C.J. Oldfield, A.M. Campen, C.M. Ratliff, K.W. Hipps, et al., Intrinsically disordered protein, *J. Mol. Graph. Model.* 19 (1) (2001) 26–59.
- [41] J.D. Durrant, J.A. McCammon, Molecular dynamics simulations and drug discovery, *BMC Biol.* 9 (1) (2011) 71.
- [42] W.A. Eaton, V. Munoz, S.J. Hagen, G.S. Jas, L.J. Lapidus, E.R. Henry, J. Hofrichter, Fast kinetics and mechanisms in protein folding, *Annu. Rev. Biophys. Biomol. Struct.* 29 (1) (2000) 327–359.
- [43] W.A. Eaton, V. Munoz, P.A. Thompson, E.R. Henry, J. Hofrichter, Kinetics and dynamics of loops,  $\alpha$ -helices,  $\beta$ -hairpins, and fast-folding proteins, *Accounts Chem. Res.* 31 (11) (1998) 745–753.
- [44] C. Edwards, M. Cohen, S. Bloom, Peptides as Drugs, *QJM* 92 (1) (1999) 1–4.
- [45] D.L. Ensign, P.M. Kasson, V.S. Pande, Heterogeneity even at the speed limit of folding: large-scale molecular dynamics study of a fast-folding variant of the villin headpiece, *J. Mol. Biol.* 374 (3) (2007) 806–816.
- [46] D.L. Ensign, V.S. Pande, The fp35 ww domain folds with structural and mechanistic heterogeneity in molecular dynamics simulations, *Biophys. J.* 96 (8) (2009) L53–L55.
- [47] J.C. Faver, M.L. Benson, X. He, B.P. Roberts, B. Wang, M.S. Marshall, C.D. Sherrill, K.M. Merz Jr., The energy computation paradox and ab initio protein folding, *PLoS One* 6 (4) (2011) e18868.
- [48] C.-J. Feng, B. Dhayalan, A. Tokmakoff, Refinement of peptide conformational ensembles by 2d ir spectroscopy: application to ala-ala-ala, *Biophys. J.* 114 (12) (2018) 2820–2832.
- [49] P. Ferrara, J. Apostolakis, A. Cafilisch, Thermodynamics and kinetics of folding of two model peptides investigated by molecular dynamics simulations, *J. Phys. Chem. B* 104 (20) (2000) 5000–5010.
- [50] P. Ferrara, A. Cafilisch, Folding simulations of a three-stranded antiparallel  $\beta$ -sheet peptide, *Proc. Natl. Acad. Sci. Unit. States Am.* 97 (20) (2000) 10780–10785.
- [51] P.L. Freddolino, C.B. Harrison, Y. Liu, K. Schulten, Challenges in protein-folding simulations, *Nat. Phys.* 6 (10) (2010) 751.
- [52] P.L. Freddolino, F. Liu, M. Gruebele, K. Schulten, Ten-microsecond molecular dynamics simulation of a fast-folding ww domain, *Biophys. J.* 94 (10) (2008) L75–L77.
- [53] Freddolino, P. L., Park, S., Roux, B., and Schulten, K. Force field bias in protein folding simulations. *Biophys. J.* 96, 9.
- [54] P.L. Freddolino, K. Schulten, Common structural transitions in explicit-solvent simulations of villin headpiece folding, *Biophys. J.* 97 (8) (2009) 2338–2347.
- [55] M.S. Friedrichs, P. Eastman, V. Vaidyanathan, M. Houston, S. Legrand, A.L. Beberg, D.L. Ensign, C.M. Bruns, V.S. Pande, Accelerating molecular dynamic simulation on graphics processing units, *J. Comput. Chem.* 30 (6) (2009) 864–872.
- [56] R. Fuentes-Azcatl, J. Alejandre, Non-polarizable force field of water based on the dielectric constant: tip4p/e, *J. Phys. Chem. B* 118 (5) (2014) 1263–1272.
- [57] A.E. Garcia, K.Y. Sanbonmatsu,  $\alpha$ -helical stabilization by side chain shielding of backbone hydrogen bonds, *Proc. Natl. Acad. Sci. Unit. States Am.* 99 (5) (2002) 2782–2787.
- [58] P.S. Georgoulia, N.M. Glykos, Using j-coupling constants for force field validation: application to hepta-alanine, *J. Phys. Chem. B* 115 (51) (2011) 15221–15227.
- [59] P.S. Georgoulia, N.M. Glykos, On the foldability of tryptophan-containing tetra- and pentapeptides: an exhaustive molecular dynamics study, *J. Phys. Chem. B* 117 (18) (2013) 5522–5532.
- [60] P.S. Georgoulia, N.M. Glykos, Folding molecular dynamics simulation of a gp41-derived peptide reconcile divergent structure determinations, *ACS Omega* 3 (11) (2018) 14746–14754.
- [61] S. Gnanakaran, A.E. Garcia, Validation of an all-atom protein force field: from dipeptides to larger peptides, *J. Phys. Chem. B* 107 (46) (2003) 12555–12557.
- [62] S. Gnanakaran, A.E. Garcia, Helix-coil transition of alanine peptides in water: force field dependence on the folded and unfolded structures, *Proteins: Struct. Func. Bioinf.* 59 (4) (2005) 773–782.
- [63] S. Gnanakaran, H. Nymeyer, J. Portman, K.Y. Sanbonmatsu, A.E. Garcia, Peptide folding simulations, *Curr. Opin. Struct. Biol.* 13 (2) (2003) 168–174.
- [64] B. Guillot, A reappraisal of what we have learnt during three decades of computer simulations on water, *J. Mol. Liq.* 101 (2002) 1–3 219–260.
- [65] O. Guvench, A.D. MacKerell, Comparison of protein force fields for molecular dynamics simulations, *Molecular Modeling of Proteins*, Springer, 2008, pp. 63–88.
- [66] S.J. Hagen, J. Hofrichter, A. Szabo, W.A. Eaton, Diffusion-limited contact formation in unfolded cytochrome c: estimating the maximum rate of protein folding, *Proc. Natl. Acad. Sci. Unit. States Am.* 93 (21) (1996) 11615–11617.
- [67] T.A. Halgren, W. Damm, Polarizable force fields, *Curr. Opin. Struct. Biol.* 11 (2) (2001) 236–242.
- [68] D. Hamelberg, J. Mongan, J.A. McCammon, Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules, *J. Chem. Phys.* 120 (24) (2004) 11919–11929.
- [69] E. Harder, W. Damm, J. Maple, C. Wu, M. Reboul, J.Y. Xiang, L. Wang, D. Luyban, M.K. Dahlgren, J.L. Knight, et al., Opls3: a force field providing broad coverage of drug-like small molecules and proteins, *J. Chem. Theor. Comput.* 12 (1) (2015) 281–296.
- [70] M.P. Hatfield, R.F. Murphy, S. Lovas, Vcd spectroscopic properties of the  $\beta$ -hairpin forming miniprotein cln025 in various solvents, *Biopolymers* 93 (5) (2010) 442–450.
- [71] A.J. Hazel, E.T. Walters, C.N. Rowley, J.C. Gumbart, Folding free energy landscapes of  $\beta$ -sheets with non-polarizable and polarizable charmm force fields, *J. Chem. Phys.* 149 (2018) 0723177.
- [72] J. Henriques, C. Cragnell, M. Skepe, Molecular dynamics simulations of intrinsically disordered proteins: force field evaluation and comparison with experiment, *J. Chem. Theor. Comput.* 11 (7) (2015) 3420–3431.

- [73] B. Hess, N.F. van der Vegt, Hydration thermodynamic properties of amino acid analogues: a systematic comparison of biomolecular force fields and water models, *J. Phys. Chem. B* 110 (35) (2006) 17616–17626.
- [74] J. Higo, N. Ito, M. Kuroda, S. Ono, N. Nakajima, H. Nakamura, Energy landscape of a peptide consisting of  $\alpha$ -helix, 310-helix,  $\beta$ -turn,  $\beta$ -hairpin, and other disordered conformations, *Protein Sci.* 10 (6) (2001) 1160–1171.
- [75] B.K. Ho, K.A. Dill, Folding very short peptides using molecular dynamics, *PLoS Comput. Biol.* 2 (4) (2006) e27.
- [76] M. Hoang Viet, P. Derreumaux, P.H. Nguyen, Communication: Multiple Atomistic Force Fields in a Single Enhanced Sampling Simulation, (2015).
- [77] S. Honda, K. Yamasaki, Y. Sawada, H. Morii, 10 residue folded peptide designed by segment statistics, *Structure* 12 (8) (2004) 1507–1518.
- [78] H.W. Horn, W.C. Swope, J.W. Pitera, J.D. Madura, T.J. Dick, G.L. Hura, T. Head-Gordon, Development of an improved four-site water model for biomolecular simulations: tip4p-ew, *J. Chem. Phys.* 120 (20) (2004) 9665–9678.
- [79] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, Comparison of multiple amber force fields and development of improved protein backbone parameters, *Proteins: Struct. Func. Bioinf.* 65 (3) (2006) 712–725.
- [80] J. Huang, A.D. MacKerell, Force field development and simulations of intrinsically disordered proteins, *Curr. Opin. Struct. Biol.* 48 (2018) 40–48.
- [81] J. Huang, A.D. MacKerell Jr., Charmm36 all-atom additive protein force field: validation based on comparison to nmr data, *J. Comput. Chem.* 34 (25) (2013) 2135–2145.
- [82] J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B.L. de Groot, H. Grubmüller, A.D. MacKerell Jr., Charmm36m: an improved force field for folded and intrinsically disordered proteins, *Nat. Methods* 14 (1) (2016) 71.
- [83] W.L. Jorgensen, Quantum and statistical mechanical studies of liquids. 10. transferable intermolecular potential functions for water, alcohols, and ethers. application to liquid water, *J. Am. Chem. Soc.* 103 (2) (1981) 335–340.
- [84] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, Comparison of simple potential functions for simulating liquid water, *J. Chem. Phys.* 79 (2) (1983) 926–935.
- [85] 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 (45) (1996) 11225–11236.
- [86] W.L. Jorgensen, J. Tirado-Rives, The opls [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin, *J. Am. Chem. Soc.* 110 (6) (1988) 1657–1666.
- [87] G.A. Kaminski, R.A. Friesner, J. Tirado-Rives, W.L. Jorgensen, Evaluation and reparameterization of the opls-aa force field for proteins via comparison with accurate quantum chemical calculations on peptides, *J. Phys. Chem. B* 105 (28) (2001) 6474–6487.
- [88] N. Kamiya, J. Higo, H. Nakamura, Conformational transition states of a  $\beta$ -hairpin peptide between the ordered and disordered conformations in explicit water, *Protein Sci.* 11 (10) (2002) 2297–2307.
- [89] T.H. Keller, A. Pichota, Z. Yin, A practical view of druggability, *Curr. Opin. Chem. Biol.* 10 (4) (2006) 357–361.
- [90] B.L. Kier, N.H. Andersen, Probing the lower size limit for protein-like fold stability: ten-residue microproteins with specific, rigid structures in water, *J. Am. Chem. Soc.* 130 (44) (2008) 14675–14683.
- [91] Y. Kliger, Computational approaches to therapeutic peptide discovery, *Pept. Sci.* 94 (6) (2010) 701–710.
- [92] A.N. Koller, H. Schwalbe, H. Gohlke, Starting structure dependence of nmr order parameters derived from md simulations: implications for judging force-field quality, *Biophys. J.* 95 (1) (2008) L04–L06.
- [93] P. Kollman, R. Dixon, W. Cornell, T. Fox, C. Chipot, A. Pohorille, The development/application of a minimalist organic/biochemical molecular mechanic force field using a combination of ab initio calculations and experimental data, *Computer Simulation of Biomolecular Systems*, Springer, 1997, pp. 83–96.
- [94] P.I. Koukos, N.M. Glykos, Folding molecular dynamics simulations accurately predict the effect of mutations on the stability and structure of a vamin-derived peptide, *J. Phys. Chem. B* 118 (34) (2014) 10076–10084.
- [95] J. Kubelka, E.R. Henry, T. Cellmer, J. Hofrichter, W.A. Eaton, Chemical, physical, and theoretical kinetics of an ultrafast folding protein, *Proc. Natl. Acad. Sci. Unit. States Am.* 105 (48) (2008) 18655–18662.
- [96] J. Kubelka, J. Hofrichter, W.A. Eaton, The protein folding speed limit, *Curr. Opin. Struct. Biol.* 14 (1) (2004) 76–88.
- [97] B. Kuhlman, D. Baker, Exploring folding free energy landscapes using computational protein design, *Curr. Opin. Struct. Biol.* 14 (1) (2004) 89–95.
- [98] P. Kührová, A. De Simone, M. Otyepka, R.B. Best, Force-field dependence of chignolin folding and misfolding: comparison with experiment and redesign, *Biophys. J.* 102 (8) (2012) 1897–1906.
- [99] N.L. Fawzi, A.H. Phillips, J.Z. Ruscio, M. Doucleff, D.E. Wemmer, T. Head-Gordon, Structure and dynamics of the  $\alpha$ 21–30 peptide from the interplay of nmr experiments and molecular simulations, *J. Am. Chem. Soc.* 130 (19) (2008) 6145–6158.
- [100] A. Laio, M. Parrinello, Escaping free-energy minima, *Proc. Natl. Acad. Sci. Unit. States Am.* 99 (20) (2002) 12562–12566.
- [101] G. Lamoureux, A.D. MacKerell Jr., B. Roux, A simple polarizable model of water based on classical drude oscillators, *J. Chem. Phys.* 119 (10) (2003) 5185–5197.
- [102] O.F. Lange, D. Van der Spoel, B.L. De Groot, Scrutinizing molecular mechanics force fields on the submicrosecond timescale with nmr data, *Biophys. J.* 99 (2) (2010) 647–655.
- [103] L.J. Lapidus, W.A. Eaton, J. Hofrichter, Measuring the rate of intramolecular contact formation in polypeptides, *Proc. Natl. Acad. Sci. Unit. States Am.* 97 (13) (2000) 7220–7225.
- [104] S.M. Larson, C.D. Snow, M. Shirts, V.S. Pande, Folding@ Home and Genome@ Home: Using Distributed Computing to Tackle Previously Intractable Problems in Computational Biology, (2009) arXiv preprint arXiv:0901.0866.
- [105] D.-W. Li, R. Brüschweiler, Nmr-based protein potentials, *Angew. Chem. Int. Ed.* 49 (38) (2010) 6778–6780.
- [106] D.-W. Li, R. Brüschweiler, Iterative optimization of molecular mechanics force fields from nmr data of full-length proteins, *J. Chem. Theor. Comput.* 7 (6) (2011) 1773–1782.
- [107] K. Lindorff-Larsen, P. Maragakis, S. Piana, M.P. Eastwood, R.O. Dror, D.E. Shaw, Systematic validation of protein force fields against experimental data, *PLoS One* 7 (2) (2012) e32131.
- [108] K. Lindorff-Larsen, S. Piana, R.O. Dror, D.E. Shaw, How fast-folding proteins fold, *Science* 334 (6055) (2011) 517–520.
- [109] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J.L. Klepeis, R.O. Dror, D.E. Shaw, Improved side-chain torsion potentials for the amber ff99sb protein force field, *Proteins: Struct. Func. Bioinf.* 78 (8) (2010) 1950–1958.
- [110] A. Loffet, Peptides as drugs: is there a market? *J. Pept. Sci.: Offic. Publ. Eur. Pept. Soc.* 8 (1) (2002) 1–7.
- [111] P.E. Lopes, O. Guvench, A.D. MacKerell, Current status of protein force fields for molecular dynamics simulations, *Molecular Modeling of Proteins*, Springer, 2015, pp. 47–71.
- [112] A.D. MacKerell Jr., Empirical force fields for biological macromolecules: overview and issues, *J. Comput. Chem.* 25 (13) (2004) 1584–1604.
- [113] A.D. MacKerell Jr., D. Bashford, M. Bellott, R.L. Dunbrack Jr., J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, et al., *J. Phys. Chem. B* 102 (18) (1998) 3586–3616.
- [114] 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, *Encycl. Comput. Chem.* 1 (2002).
- [115] A.D. MacKerell Jr., M. Feig, C.L. Brooks, Improved treatment of the protein backbone in empirical force fields, *J. Am. Chem. Soc.* 126 (3) (2003) 698–699.
- [116] 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 (11) (2004) 1400–1415.
- [117] M.W. Mahoney, W.L. Jorgensen, A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions, *J. Chem. Phys.* 112 (20) (2000) 8910–8922.
- [118] J.A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K.E. Hauser, C. Simmerling, ff14sb: improving the accuracy of protein side chain and backbone parameters from ff99sb, *J. Chem. Theor. Comput.* 11 (8) (2015) 3696–3713.
- [119] V.H. Man, P.H. Nguyen, P. Derreumaux, High-resolution structures of the amyloid- $\beta$  1–42 dimers from the comparison of four atomistic force fields, *J. Phys. Chem. B* 121 (24) (2017) 5977–5987.
- [120] E. Marinari, G. Parisi, Simulated tempering: a new Monte Carlo scheme, *EPL (Europhys. Lett.)* 19 (6) (1992) 451.
- [121] P. Mark, L. Nilsson, Structure and dynamics of the tip3p, spc, and spc/e water models at 298 k, *J. Phys. Chem.* 105 (43) (2001) 9954–9960.
- [122] P.R. Markwick, J.A. McCammon, Studying functional dynamics in bio-molecules using accelerated molecular dynamics, *Phys. Chem. Chem. Phys.* 13 (45) (2011) 20053–20065.
- [123] Matthes, D., and De Groot, B. L. Secondary structure propensities in peptide folding simulations: a systematic comparison of molecular mechanics interaction schemes.
- [124] J. Maupetit, P. Derreumaux, P. Tuffery, Pep-fold: an online resource for de novo peptide structure prediction, *Nucleic Acids Res.* 37 (suppl.2) (2009) W498–W503.
- [125] U. Mayor, C.M. Johnson, V. Daggett, A.R. Fersht, Protein folding and unfolding in microseconds to nanoseconds by experiment and simulation, *Proc. Natl. Acad. Sci. Unit. States Am.* 97 (25) (2000) 13518–13522.
- [126] J.A. McCammon, A speed limit for protein folding, *Proc. Natl. Acad. Sci. U.S.A.* 93 (21) (1996) 11426.
- [127] K.A. McKiernan, B.E. Husic, V.S. Pande, Modeling the mechanism of cln025  $\beta$ -hairpin formation, *J. Chem. Phys.* 147 (10) (2017) 104107.
- [128] D. Mercadante, S. Milles, G. Fuentes, D.I. Svergun, E.A. Lemke, F. Grater, Kirkwood–buff approach rescues overcollapse of a disordered protein in canonical protein force fields, *J. Phys. Chem. B* 119 (25) (2015) 7975–7984.
- [129] D. Mercadante, J.A. Wagner, I.V. Aramburu, E.A. Lemke, F. Grater, Sampling long-versus short-range interactions defines the ability of force fields to reproduce the dynamics of intrinsically disordered proteins, *J. Chem. Theor. Comput.* 13 (9) (2017) 3964–3974.
- [130] A. Mitsutake, Y. Sugita, Y. Okamoto, Generalized-ensemble algorithms for molecular simulations of biopolymers, *Pept. Sci.: Orig. Res. Biomol.* 60 (2) (2001) 96–123.
- [131] J. Mittal, R.B. Best, Tackling force-field bias in protein folding simulations: folding of villin hp35 and pin ww domains in explicit water, *Biophys. J.* 99 (3) (2010) L26–L28.
- [132] V. Munoz, P.A. Thompson, J. Hofrichter, W.A. Eaton, Folding dynamics and mechanism of  $\beta$ -hairpin formation, *Nature* 390 (6656) (1997) 196.
- [133] J. Nasic-Labouze, P.H. Nguyen, F. Sterpone, O. Berthoumiou, N.-V. Buchete, S. Cote, A. De Simone, A.J. Doig, P. Fallar, A. Garcia, et al., Amyloid  $\beta$  protein and alzheimer's disease: when computer simulations complement experimental studies, *Chem. Rev.* 115 (9) (2015) 3518–3563.
- [134] E. Neria, S. Fischer, M. Karplus, Simulation of activation free energies in molecular systems, *J. Chem. Phys.* 105 (5) (1996) 1902–1921.
- [135] D. Nettelts, S. Müller-Späh, F. Küster, H. Hoffmann, D. Haenni, S. Rügger, L. Reymond, A. Hoffmann, J. Kubelka, B. Heinz, et al., Single-molecule

- spectroscopy of the temperature-induced collapse of unfolded proteins, *Proc. Natl. Acad. Sci. Unit. States Am.* 106 (49) (2009) 20740–20745.
- [136] P.H. Nguyen, M.S. Li, P. Derreumaux, Effects of all-atom force fields on amyloid oligomerization: replica exchange molecular dynamics simulations of the  $\alpha\beta$  16–22 dimer and trimer, *Phys. Chem. Chem. Phys.* 13 (20) (2011) 9778–9788.
- [137] A. Okur, B. Strockbine, V. Hornak, C. Simmerling, Using pc clusters to evaluate the transferability of molecular mechanics force fields for proteins, *J. Comput. Chem.* 24 (1) (2003) 21–31.
- [138] S. Ono, N. Nakajima, J. Higo, H. Nakamura, Peptide free-energy profile is strongly dependent on the force field: comparison of c96 and amber95, *J. Comput. Chem.* 21 (9) (2000) 748–762.
- [139] A.V. Onufriev, S. Izadi, *Water models for biomolecular simulations*, Wiley Interdiscipl. Rev.: Comput. Mol. Sci. 8 (2) (2018) e1347.
- [140] 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.
- [141] K. Ostermeir, M. Zacharias, Advanced replica-exchange sampling to study the flexibility and plasticity of peptides and proteins, *Biochim. Biophys. Acta Protein Proteomics* 5 (2013) (1834) 847–853.
- [142] J.F. Ouyang, R. Bettens, Modelling water: a lifetime enigma, *CHIMIA Int. J. Chem.* 69 (3) (2015) 104–111.
- [143] V.S. Pande, D.S. Rokhsar, Molecular dynamics simulations of unfolding and re-folding of a  $\beta$ -hairpin fragment of protein g, *Proc. Natl. Acad. Sci. Unit. States Am.* 96 (16) (1999) 9062–9067.
- [144] K.K. Patapati, N.M. Glykos, Three force fields' views of the 310 helix, *Biophys. J.* 101 (7) (2011) 1766–1771.
- [145] I. Patmanidis, N.M. Glykos, As good as it gets? folding molecular dynamics simulations of the lyta choline-binding peptide result to an exceptionally accurate model of the peptide structure, *J. Mol. Graph. Model.* 41 (2013) 68–71.
- [146] A. Perez, J.L. MacCallum, E. Brini, C. Simmerling, K.A. Dill, Grid-based backbone correction to the ff12sb protein force field for implicit-solvent simulations, *J. Chem. Theor. Comput.* 11 (10) (2015) 4770–4779.
- [147] J.C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R.D. Skeel, L. Kale, K. Schulten, Scalable molecular dynamics with namd, *J. Comput. Chem.* 26 (16) (2005) 1781–1802.
- [148] S. Piana, A.G. Donchev, P. Robustelli, D.E. Shaw, Water dispersion interactions strongly influence simulated structural properties of disordered protein states, *J. Phys. Chem. B* 119 (16) (2015) 5113–5123.
- [149] S. Piana, J.L. Klepeis, D.E. Shaw, Assessing the accuracy of physical models used in protein-folding simulations: quantitative evidence from long molecular dynamics simulations, *Curr. Opin. Struct. Biol.* 24 (2014) 98–105.
- [150] S. Piana, K. Lindorff-Larsen, D.E. Shaw, How robust are protein folding simulations with respect to force field parameterization? *Biophys. J.* 100 (9) (2011) L47–L49.
- [151] J.W. Pitera, W. Swope, Understanding folding and design: replica-exchange simulations of "trp-cage" miniproteins, *Proc. Natl. Acad. Sci. Unit. States Am.* 100 (13) (2003) 7587–7592.
- [152] E.A. Ploetz, N. Bentein, P.E. Smith, Developing force fields from the microscopic structure of solutions, *Fluid Phase Equilib.* 290 (2010) 1–2 43–47.
- [153] J.W. Ponder, D.A. Case, Force fields for protein simulations, *Advances in Protein Chemistry*, vol. 66, Elsevier, 2003, pp. 27–85.
- [154] S. Rauscher, V. Gapsys, M.J. Gajda, M. Zweckstetter, B.L. de Groot, H. Grubmüller, Structural ensembles of intrinsically disordered proteins depend strongly on force field: a comparison to experiment, *J. Chem. Theor. Comput.* 11 (11) (2015) 5513–5524.
- [155] M.M. Reif, P.H. Hunenberger, C. Oostenbrink, New interaction parameters for charged amino acid side chains in the gromos force field, *J. Chem. Theor. Comput.* 8 (10) (2012) 3705–3723.
- [156] P. Robustelli, S. Piana, D.E. Shaw, Developing a molecular dynamics force field for both folded and disordered protein states, *Proc. Natl. Acad. Sci. Unit. States Am.* 115 (21) (2018) 20180690E4758–E4766.
- [157] A. Rodriguez, P. Mokoema, F. Corcho, K. Bissett, J.J. Perez, Computational study of the free energy landscape of the miniprotein cln025 in explicit and implicit solvent, *J. Phys. Chem. B* 115 (6) (2011) 1440–1449.
- [158] D.J. Rosenman, C. Wang, A.E. Garcia, Characterization of  $\alpha\beta$  monomers through the convergence of ensemble properties among simulations with multiple force fields, *J. Phys. Chem. B* 120 (2) (2015) 259–277.
- [159] R. Salomon-Ferrer, A.W. Gotz, D. Poole, S. Le Grand, R.C. Walker, Routine microsecond molecular dynamics simulations with amber on gpus. 2. explicit solvent particle mesh ewald, *J. Chem. Theor. Comput.* 9 (9) (2013) 3878–3888.
- [160] D. Satoh, K. Shimizu, S. Nakamura, T. Terada, Folding free-energy landscape of a 10-residue mini-protein, chignolin, *FEBS Lett.* 580 (14) (2006) 3422–3426.
- [161] N. Schmid, A.P. Eichenberger, A. Choutko, S. Riniker, M. Winger, A.E. Mark, W.F. van Gunsteren, Definition and testing of the gromos force-field versions 54a7 and 54b7, *Eur. Biophys. J.* 40 (7) (2011) 843.
- [162] A.-P. Serafeim, G. Salamanos, K.K. Patapati, N.M. Glykos, Sensitivity of folding molecular dynamics simulations to even minor force field changes, *J. Chem. Inf. Model.* 56 (10) (2016) 2035–2041.
- [163] D.E. Shaw, R.O. Dror, J.K. Salmon, J. Grossman, K.M. Mackenzie, J.A. Bank, C. Young, M.M. Deneroff, B. Batson, K.J. Bowers, et al., Millisecond-scale molecular dynamics simulations on anton, *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis*, ACM, 2009, p. 39.
- [164] D.E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R.O. Dror, M.P. Eastwood, J.A. Bank, J.M. Jumper, J.K. Salmon, Y. Shan, et al., Atomic-level characterization of the structural dynamics of proteins, *Science* 330 (6002) (2010) 341–346.
- [165] M.S. Shell, S.B. Ozkan, V. Voelz, G.A. Wu, K.A. Dill, Blind test of physics-based prediction of protein structures, *Biophys. J.* 96 (3) (2009) 917–924.
- [166] Y. Shi, Z. Xia, J. Zhang, R. Best, C. Wu, J.W. Ponder, P. Ren, Polarizable atomic multipole-based amoeba force field for proteins, *J. Chem. Theor. Comput.* 9 (9) (2013) 4046–4063.
- [167] M. Shirts, V.S. Pande, Screen savers of the world unite!, *Science* 290 (5498) (2000) 1903–1904.
- [168] S.A. Showalter, R. Brüschweiler, Quantitative molecular ensemble interpretation of nmr dipolar couplings without restraints, *J. Am. Chem. Soc.* 129 (14) (2007) 4158–4159.
- [169] S.A. Showalter, E. Johnson, M. Rance, R. Brüschweiler, Toward quantitative interpretation of methyl side-chain dynamics from nmr by molecular dynamics simulations, *J. Am. Chem. Soc.* 129 (46) (2007) 14146–14147.
- [170] C. Simmerling, B. Strockbine, A.E. Roitberg, All-atom structure prediction and folding simulations of a stable protein, *J. Am. Chem. Soc.* 124 (38) (2002) 11258–11259.
- [171] C.L. Simmerling, R. Elber, Computer determination of peptide conformations in water: different roads to structure, *Proc. Natl. Acad. Sci. Unit. States Am.* 92 (8) (1995) 3190–3193.
- [172] K.T. Simons, C. Kooperberg, E. Huang, D. Baker, Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and bayesian scoring functions, *J. Mol. Biol.* 268 (1) (1997) 209–225.
- [173] C.M. Siwy, C. Lockhart, D.K. Klimov, Is the conformational ensemble of alzheimer's  $\alpha$ 10–40 peptide force field dependent? *PLoS Comput. Biol.* 13 (2017) e10053141.
- [174] M.D. Smith, J.S. Rao, E. Segelken, L. Cruz, Force-field induced bias in the structure of  $\alpha$ 21–30: a comparison of opls, amber, charmm, and gromos force fields, *J. Chem. Inf. Model.* 55 (12) (2015) 2587–2595.
- [175] C.D. Snow, H. Nguyen, V.S. Pande, M. Gruebele, Absolute comparison of simulated and experimental protein-folding dynamics, *Nature* 420 (6911) (2002) 102.
- [176] A.K. Somavarapu, K.P. Kepp, The dependence of amyloid- $\beta$  dynamics on protein force fields and water models, *ChemPhysChem* 16 (15) (2015) 3278–3289.
- [177] E.J. Sorin, V.S. Pande, Exploring the helix-coil transition via all-atom equilibrium ensemble simulations, *Biophys. J.* 88 (4) (2005) 2472–2493.
- [178] P.J. Steinbach, Exploring peptide energy landscapes: a test of force fields and implicit solvent models, *Proteins: Struct. Funct. Bioinf.* 57 (4) (2004) 665–677.
- [179] G. Stirnemann, F. Sterpone, Recovering protein thermal stability using all-atom Hamiltonian replica-exchange simulations in explicit solvent, *J. Chem. Theor. Comput.* 11 (12) (2015) 5573–5577.
- [180] J.E. Stone, D.J. Hardy, I.S. Ufimtsev, K. Schulten, Gpu-accelerated molecular modeling coming of age, *J. Mol. Graph. Model.* 29 (2) (2010) 116–125.
- [181] Y. Sugita, Y. Okamoto, Replica-exchange molecular dynamics method for protein folding, *Chem. Phys. Lett.* 314 (1–2) (1999) 141–151.
- [182] G.M. Torrie, J.P. Valleau, Nonphysical sampling distributions in Monte Carlo free-energy estimation: umbrella sampling, *J. Comput. Phys.* 23 (2) (1977) 187–199.
- [183] L. Tran, T. Ha-Duong, Exploring the alzheimer amyloid- $\beta$  peptide conformational ensemble: a review of molecular dynamics approaches, *Peptides* 69 (2015) 86–91.
- [184] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A.E. Mark, H.J. Berendsen, Gromacs: fast, flexible, and free, *J. Comput. Chem.* 26 (16) (2005) 1701–1718.
- [185] D. van der Spoel, P.J. van Maaren, H.J. Berendsen, A systematic study of water models for molecular simulation: derivation of water models optimized for use with a reaction field, *J. Chem. Phys.* 108 (24) (1998) 10220–10230.
- [186] van Gunsteren, W. F., Billeter, S. R., Eising, A. A., Hünenberger, P. H., Krüger, P., Mark, A. E., Scott, W. R., and Tironi, I. G. *Biomolecular Simulation: the {GROMOS96} Manual and User Guide*.
- [187] K. Vanommeslaeghe, A. MacKerell, Charmm additive and polarizable force fields for biophysics and computer-aided drug design, *Biochim. Biophys. Acta Gen. Subj.* 5 (1850) 861–871 2015.
- [188] P. Vlieghe, V. Lisowski, J. Martinez, M. Khrestchatsky, Synthetic therapeutic peptides: science and market, *Drug Discov. Today* 15 (2010) 1–2 40–56.
- [189] A. Wallqvist, R.D. Mountain, Molecular models of water: derivation and description, *Rev. Comput. Chem.* (1999) 183–247.
- [190] J. Wang, P. Cieplak, P.A. Kollman, How well does a restrained electrostatic potential (resp) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* 21 (12) (2000) 1049–1074.
- [191] L.-P. Wang, T.J. Martinez, V.S. Pande, Building force fields: an automatic, systematic, and reproducible approach, *J. Phys. Chem. Lett.* 5 (11) (2014) 1885–1891.
- [192] L.-P. Wang, K.A. McKiernan, J. Gomes, K.A. Beauchamp, T. Head-Gordon, J.E. Rice, W.C. Swope, T.J. Martínez, V.S. Pande, Building a more predictive protein force field: a systematic and reproducible route to amber-fb15, *J. Phys. Chem. B* 121 (16) (2017) 4023–4039.
- [193] W. Wang, W. Ye, C. Jiang, R. Luo, H.-F. Chen, New force field on modeling intrinsically disordered proteins, *Chem. Biol. Drug Des.* 84 (3) (2014) 253–269.
- [194] C.-C. Wei, M.-H. Ho, W.-H. Wang, Y.-C. Sun, Molecular dynamics simulation of folding of a short helical peptide with many charged residues, *J. Phys. Chem. B* 109 (42) (2005) 19980–19986.
- [195] S.J. Weiner, P.A. Kollman, D.A. Case, U.C. Singh, C. Ghio, G. Alagona, S. Profeta, P. Weiner, A new force field for molecular mechanical simulation of nucleic acids and proteins, *J. Am. Chem. Soc.* 106 (3) (1984) 765–784.
- [196] S.J. Weiner, P.A. Kollman, D.T. Nguyen, D.A. Case, An all atom force field for simulations of proteins and nucleic acids, *J. Comput. Chem.* 7 (2) (1986) 230–252.
- [197] L. Wickstrom, A. Okur, C. Simmerling, Evaluating the performance of the ff99sb force field based on nmr scalar coupling data, *Biophys. J.* 97 (3) (2009) 853–856.
- [198] L. Wickstrom, A. Okur, K. Song, V. Hornak, D.P. Raleigh, C.L. Simmerling, The unfolded state of the villin headpiece helical subdomain: computational studies of the role of locally stabilized structure, *J. Mol. Biol.* 360 (5) (2006) 1094–1107.

- [199] S. Williams, T.P. Causgrove, R. Gilmanshin, K.S. Fang, R.H. Callender, W.H. Woodruff, R.B. Dyer, Fast events in protein folding: helix melting and formation in a small peptide, *Biochemistry* 35 (3) (1996) 691–697.
- [200] P.E. Wright, H.J. Dyson, R.A. Lerner, Conformation of peptide fragments of proteins in aqueous solution: implications for initiation of protein folding, *Biochemistry* 27 (19) (1988) 7167–7175.
- [201] L. Yang, C.-h. Tan, M.-J. Hsieh, J. Wang, Y. Duan, P. Cieplak, J. Caldwell, P.A. Kollman, R. Luo, New-generation amber united-atom force field, *J. Phys. Chem. B* 110 (26) (2006) 13166–13176.
- [202] W.Y. Yang, M. Gruebele, Folding at the speed limit, *Nature* 423 (6936) (2003) 193.
- [203] W. Ye, D. Ji, W. Wang, R. Luo, H.-F. Chen, Test and evaluation of ff99idps force field for intrinsically disordered proteins, *J. Chem. Inf. Model.* 55 (5) (2015) 1021–1029.
- [204] T. Yoda, Y. Sugita, Y. Okamoto, Comparisons of force fields for proteins by generalized-ensemble simulations, *Chem. Phys. Lett.* 386 (2004) 4–6 460–467.
- [205] B. Zagrovic, C.D. Snow, M.R. Shirts, V.S. Pande, Simulation of folding of a small alpha-helical protein in atomistic detail using worldwide-distributed computing, *J. Mol. Biol.* 323 (5) (2002) 927–937.
- [206] C. Zhang, J. Ma, Enhanced sampling and applications in protein folding in explicit solvent, *J. Chem. Phys.* 132 (24) (2010) 244101.
- [207] T. Zhang, P.H. Nguyen, J. Nasica-Labouze, Y. Mu, P. Derreumaux, Folding atomistic proteins in explicit solvent using simulated tempering, *J. Phys. Chem. B* 119 (23) (2015) 6941–6951.
- [208] H. Zhao, A. Caflisch, Molecular dynamics in drug design, *Eur. J. Med. Chem.* 91 (2015) 4–14.
- [209] P. Zhou, C. Wang, Y. Ren, C. Yang, F. Tian, Computational peptidology: a new and promising approach to therapeutic peptide design, *Curr. Med. Chem.* 20 (15) (2013) 1985–1996.
- [210] X. Zhu, P.E. Lopes, A.D. MacKerell Jr., Recent developments and applications of the charmm force fields, *Wiley Interdiscipl. Rev.: Comput. Mol. Sci.* 2 (1) (2012) 167–185.