Computational and Structural Biotechnology Journal 17 (2019) 1162–1170

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

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

Mini Review

Applications of Molecular Dynamics Simulation in Structure Prediction of Peptides and Proteins

Hao Geng [a,1], Fangfang Chen [b,1], Jing Ye [b], Fan Jiang [a,c,*]

[a] Lab of Computational Chemistry and Drug Design, State Key Laboratory of Chemical Oncogenomics, Peking University Shenzhen Graduate School, Shenzhen 518055, China
[b] Guangdong and Shenzhen Key Laboratory of Male Reproductive Medicine and Genetics, Peking University Shenzhen Hospital, Shenzhen PKU-HKUST Medical Center, Shenzhen 518036, China
[c] NanoAI Biotech Co.,Ltd., Silicon Valley Compound, Longhua District, Shenzhen 518109, China

ARTICLE INFO

Article history:
Received 2 April 2019
Received in revised form 7 July 2019
Accepted 23 July 2019
Available online 26 July 2019

Keywords:
Molecular dynamics
Protein folding
Cyclic peptides
Model refinement
Data-assisted modeling

ABSTRACT

Compared with rapid accumulation of protein sequences from high-throughput DNA sequencing, obtaining experimental 3D structures of proteins is still much more difficult, making protein structure prediction (PSP) potentially very useful. Currently, a vast majority of PSP efforts are based on data mining of known sequences, structures and their relationships (informatics-based). However, if closely related template is not available, these methods are usually much less reliable than experiments. They may also be problematic in predicting the structures of naturally occurring or designed peptides. On the other hand, physics-based methods including molecular dynamics (MD) can utilize our understanding of detailed atomic interactions determining biomolecular structures. In this mini-review, we show that all-atom MD can predict structures of cyclic peptides and other peptide foldamers with accuracy similar to experiments. Then, some notable successes in reproducing experimental 3D structures of small proteins through MD simulations (some with replica-exchange) of the folding were summarized. We also describe advancements of MD-based refinement of structure models, and the integration of limited experimental or bioinformatics data into MD-based structure modeling.

Contents

1. Introduction

Three-dimensional (3D) structures of proteins and their complexes provide invaluable information, not only for understanding the molecular basis of the machinery of life, but also for screening and designing of new drugs for medical applications [1]. Since the first protein 3D structure (of myoglobin) was solved by X-ray crystallography sixty years ago [2,3], enormous efforts have been paid for protein structure determinations [4–9]. However, to obtain high-resolution structure of a protein experimentally is still quite expensive and time-consuming. On the other hand, the cost of obtaining new protein sequences has dramatically decreased due to significant progresses in high-throughput DNA sequencing technology [10,11]. Therefore, there is a huge and increasing gap between the numbers of known structures and sequences. Thus, protein structure prediction (PSP) has become a cost-effective and

* Corresponding author at: Lab of Computational Chemistry and Drug Design, State Key Laboratory of Chemical Oncogenomics, Peking University Shenzhen Graduate School, Shenzhen 518055, China.
E-mail address: jiangfan@pku.edu.cn (F. Jiang).
[1] These authors contributed equally to this paper.

https://doi.org/10.1016/j.csbj.2019.07.010

high-throughput way to provide structure information for biological and pharmaceutical researches.

Various methods for PSP have been developed, which can generally be classified into two types: template-based modeling (TBM) and *de novo* structure prediction. TBM predicts the native structure of a protein target by identifying known protein structure(s) as template (s) [12], based on sequence-sequence (homology modeling) or sequence-structure (threading) alignments. Without suitable template, *de novo* PSP (template-free modeling) can be used to predict novel protein fold. The most popular and successful ones are usually based on assembly of known structure fragments with potential energy (scoring) functions from mining know protein structures (such as Rosetta [13,14] and QUARK [15]). Also, inter-residue contacts inferred from co-evolutionary signals in sequence homologs can significantly facilitate the *de novo* PSP [16,17]. The predicted protein models can span a broad range of accuracies and are potentially suitable for different applications [18].

All above popular PSP methods require certain database(s) of sequences and structures, which are thus called knowledge-based. However, the fundamental theory supporting PSP is that the 3D structure of a native-state protein in physiological condition is encoded in its amino acid sequence, as its lowest-free-energy conformation [19,20]. Thus, in theory, PSP can be achieved with only an accurate energy function with a proper conformational search method. Most popular conformational search methods include Monte Carlo (MC) and molecular dynamics (MD). MC approach has been successfully used to study the folding of peptides and proteins, using either atomistic models [21–23] or more coarse-grained models [24–28]. It has also been used to predict protein loop structures [29]. In theory, MC methods can be as accurate as MD methods, but MC may suffer from lower efficiency when using fine-grained representation of the system, especially with large number of explicit solvent molecules. There is also a danger of biasing the sampling when using MC.

MD simulation methods have a long history. The method was originally developed about sixty years ago, for theoretical physicists to study systems of many interacting particles (such as atoms or atom groups) under classical mechanics [30,31]. The now-dominant version of MD was also soon established [32], which numerically solves classical equations of motion according to physical force on each particle. Because small integration time steps of femtoseconds (10^{-15} s) are usually necessary for all-atom simulations, to simulate biologically-relevant event (such as protein folding on > microseconds, 10^{-6} s) by MD requires a huge number of numeric calculations. Over the past 40 years, the timescales that can be reached by atomistic MD simulation are increasing rapidly, even faster than the Moore's law [33,34]. Now, MD simulations have become an important and pervasive physics-based method to explore the conformational space of peptides and proteins, which can even fold small proteins (<80 amino acids) to their native structures [35].

Here we review applications of MD simulation in *ab initio* structure prediction of peptides and small proteins, refinement of protein structure models, as well as structure modeling assisted by experimental or bioinformatics data. However, we will not review following research areas traditionally *not* regarded as a PSP problem, although they are somehow related to structure prediction and heavily relying on MD simulations:

1. MD simulation has been extensively used in studying the conformational dynamics of proteins, for a long time [36]. This can be regarded as the prediction of an ensemble of structures for a protein in the native state [37]. However, this type of MD studies rely on known (representative or average) structure of a protein.
2. Some proteins or protein segments do not fold to ordered structures in the native condition [38]. The prediction of structure ensembles of intrinsically disordered proteins is very important, and MD simulation also plays a very important role [39].

Of course, we cannot cover all related works in this mini-review, but intended to give some representative examples about the topic of MD-based structure prediction of peptides and proteins.

2. Methodology Developments for MD Simulation

Since the first protein MD simulation in 1977 [36], with rapidly developing computing hardware and software, significantly longer simulation times and larger simulated systems can be achieved. Supercomputers have been built for biomolecular MD simulations such as protein folding [40]. Software packages for highly efficient MD simulations on parallel computing architectures have also been developed [41,42]. Using special-purpose supercomputer *Anton* [43], all-atom MD simulation reaching millisecond time-scale has been reported in 2010 [44]. Recently, the most exciting progress is the development of software for routine use of graphic processing unit (GPU) for MD simulations [45–47]. Now, MD simulation is becoming a powerful tool extensively used in studying biomolecular systems. Currently, a few hundred nanoseconds (ns) per day can be routinely achieved on a small protein system in explicit solvent.

In an MD simulation, forces are often calculated using a potential energy function (*force field*) of the system, which is crucial to the reliability of the MD simulation. A protein force field contains terms for both bonded (for bond lengths, bond angles, and dihedral angles) and non-bonded interactions (van der Waals and electrostatics). It has relatively simple mathematical formula, but usually contain many empirical parameters that determine its accuracy. With increasing computational performance and development of enhanced sampling methods (will be described below), more and more inaccuracies in protein force fields have been revealed [48,49].

These findings have been continuously spurring improvements of classical protein force fields, including AMBER [50–53], CHARMM [54–57], and OPLS-AA [58–60]. Most of the recent updates have been focused on the parameters of backbone and/or side-chain dihedral-angle (torsion) terms to fit *ab initio* quantum mechanics calculations or experimental (especially NMR) observables, because these parameters are closely related to the conformational behavior of simulated peptide or protein.

Water molecules plays a crucial role in driving protein folding [61], and determining the structure and dynamics of protein molecules [62]. Water models like TIP3P [63] and TIP4P [63] developed in the early 1980s have been able to accurately describe the various physicochemical properties of water at room temperature, and are still widely used in MD. In these explicit-solvent simulations, a vast majority of the computational resources are consumed in calculating forces on water molecules. To increase efficiency, methods to treat water solvent implicitly have been developed, mostly based on the Generalized Born (GB) solvation model [64–66]. Sometimes, an energy term related to solvent-accessible surface areas (SA) was used for approximating the non-polar contributions to solvation [64].

Besides the accuracy and reliability of force field (including solvent model), another important factor severely limits the applications of MD: the time scale that can be easily achieved by a conventional MD simulation is usually much shorter than those related to real problems. Thus, enhanced sampling methods have been developed. Some use biased potential to force barrier crossing on pre-defined reaction coordinates (collective variables), such as umbrella sampling [67] and metadynamics [68]. However, these methods can hardly be used in actual structure prediction because the end point of folding should be unknown. Thus, enhanced sampling methods using energy as a reaction coordinate were developed, including biased potential methods such as accelerated molecular dynamics [69], generalized ensemble methods such as replica-exchange molecular dynamics (REMD) [70], and methods combining the two aspects [71]. Currently, the REMD method becomes the most popular one for *ab initio* folding, partly because it

can utilize current main-stream multi-node parallel computing architectures very efficiently. In addition, replica-exchange method can also be used with MC simulation, which has been used in I-TASSER [72] structure prediction pipeline by Zhang's group to facilitate the fragment assembly and conformational search.

3. Peptide Structure Prediction

Peptides fill the gap between small-molecule drugs and proteins, and potentially can have both their advantages [73]. Numerous peptides have been found in natural products, and some of them have become drugs with great biological functional diversity. Still, millions of peptide sequences are estimated occurring in prokaryotic genomes [74], plants [75], venom in eukaryotes [76,77] and there are even larger sequence space for designed peptides as chemical tools and drug leads [78–81]. Because the function of a peptides is always related to its unique conformational behavior [82,83], accurate peptide structure prediction (PepSP) would contribute significantly to the peptide-based drug design.

Many attempts have been made for developing PepSP, including evaluations of some common PSP methods (Rosetta, I-TASSER) and specific development of PepSP methods (PepLook, Pep-Fold). However, these methods often cannot consistently predict the experimental structures of peptides. For example, Rosetta cannot reproduce the experimental structure of an α-conotoxin [84]. Peplook, Pep-Fold and I-Tasser predicted a set of 38 cyclic peptides consisting of 5-30 residues with average backbone RMSD (root-mean-square deviation from corresponding experimental structure) values of 3.8 Å, 4.1Å and 2.5 Å, respectively [85]. At the same time, solvent effect has been found very important for conformations of peptides [83,86], which can be treated explicitly in MD simulation, and the configuration entropy can be considered. In addition, a peptide in solution can have distinct conformations [87,88] with small free energy difference, bringing high demand on the accuracy of the energy function and solvent model.

For many naturally occurring or designed peptides, MD-based methods can be very suitable for predicting their 3D structures. Firstly, alignment of short sequences may be less reliable, hindering the use of comparative modeling techniques for PepSP [89]. Secondly, structures of short peptides are usually highly sensitive to their exact sequences, small variations may cause massive conformational alterations [90]. Third, compared with proteins, converged conformational sampling of small peptides can be much more easily achieved by MD simulations, especially with enhanced sampling methods.

Dill and coworkers studied the structure prediction of peptide fragments using REMD simulations with implicit solvent [91,92]. For 133 8-residue fragments from six different proteins, 85 of them have no preferred structure, while the structures of 41 out of 48 structured peptides bear some resemblance to their native structures [91]. Most linear peptides do not have a stable structure that can be determined by experiments, making it very difficult to benchmark PepSP methods on them. Still, there are some designed foldable linear peptides, including β-peptides [93,94], that can be folded by MD simulations two decades ago.

Cyclization of linear peptides, either by backbone or side-chain linkages, can achieve much lower conformational flexibility [95,96] and better drug-like properties, which are attracting more and more attentions. Because availability of their experimental structures (especially X-ray crystal structures) and resemblance to protein loops, we think that cyclic peptides (CPs) are much better candidate for benchmarking structure prediction methods, compared with linear peptides.

REMD simulations using AMBER ff96 with GBSA implicit solvent were performed to 18 cyclic RGD peptides [97], showing agreements between the simulated and experimental inter-proton distances. Among them, the predicted structure of cilengitide (5 residues) have 0.25 Å backbone-RMSD to its crystal structure bound to integrin. A structure prediction study of peptoid [98] shows, REMD simulation using generalized AMBER force field (GAFF) [99] with implicit solvation, combing a quantum mechanical refinement, correctly predicted the X-ray crystallographic structures of a N-aryl peptoid trimer and a cyclic peptoid nonamer to the accuracy of 0.2 Å and 1.0 Å RMSDs, respectively.

Recently, MD simulations using explicit solvent models have become more and more widely used. REMD simulation of Vasopressin and Oxytocin (peptide hormones cyclized by disulfide bond) and their mutants were performed [100], and the resulted structural ensembles were validated against experimental NMR chemical shifts and $^{3}J_{H\alpha HN}$ scalar couplings. In 2016, Lin and coworkers investigated the ability of MD with bias-exchange metadynamics (BE-META) to predict the structure of a cyclic octapeptide [101]. Representative structures of five most commonly populated states from simulations with four different force fields gave RMSD ranging from 1.11 Å to 1.88 Å to the NMR structure.

Recently, we simulated 20 all-trans CPs of 5-12 residues using explicit-water REMD with four different force fields: AMBER ff99SB – ILDN, OPLS-AA/L, RSFF1, and RSFF2 [102,103]. Our recently developed RSFF2 [104] force field performs the best (Fig. 3), which can predict the crystal structures of 17 out of these 20 CPs with backbone RMSD < 1.1 Å, and 8 CPs with backbone RMSD < 0.5Å. Metadynamics with RSFF2 was used recently by Lin and coworkers to study sequence–structure relationships of some simple cyclic hexapeptides [90], and design well-structured cyclic pentapeptides [105]. REMD with RSFF2 has also been used to study α-helical stapled peptides, together with GAFF to describe the non-standard amino acids, giving prediction in excellent agreements with experimental observations [106–108]. In one case with X-ray crystal structure solved, the predicted structure is almost identical to experimental one (RMSD of 0.3Å). (See Figs. 1 and 2).

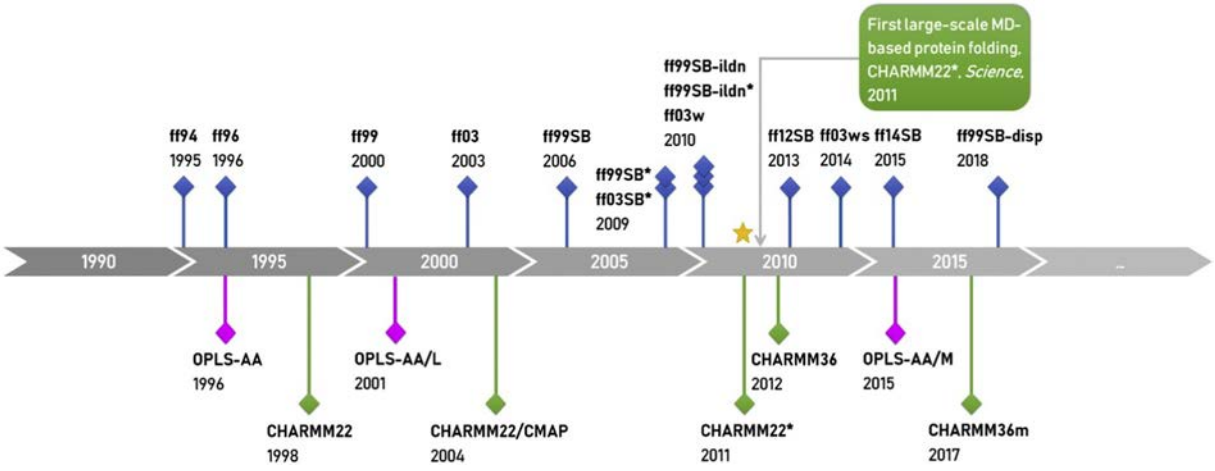

Fig. 1. Developments of the variants of the three most popular all-atom protein force fields: AMBER, CHARMM, and OPLS-AA.

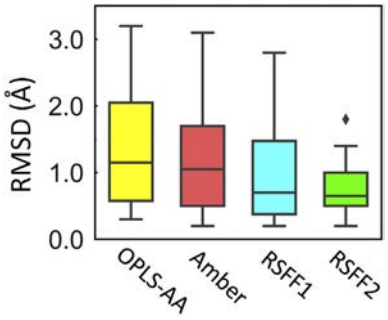

Fig. 2. Statistics (box plots) of the backbone RMSD values between predicted structures from REMD simulations with various force fields and their X-ray crystal structures of 20 all-trans CPs of 5-12 residues.

However, in spite of good conformational sampling and advanced force fields, MD simulation still cannot accurately predict peptide structures in all cases. Very recently, it was shown that the *cis*/*trans* isomerization of N-methylated cyclic hexapeptides cannot be reliably predicted [109]. Recently, simulation of guanylin (a 15-residue peptide with four cysteines) showed that the distribution of three disulfide-bond isomers is in qualitative agreement with experiment, but the most stable conformation of the isomer 2 is significantly different from the poorly ordered structure of the truncated peptide [110]. In another study, BE-META simulations were performed on five cyclic isoDGR-containing α/β-peptides using eight widely used force fields with explicit water to reproduce 79 NMR observables [111]. Most of the force fields display good agreement with experimental $^3J_{(HN,H\alpha)}$, but poor agreement was observed for NMR observables directly related to β-amino acids. In addition, accelerated MD simulations of three CPs were performed to predict their conformations [112], and benchmarked against inter-proton distances from NMR experiments and X-ray structures, revealing discrepancy between calculations and experimental observations. These studies are showing that MD-based PepSP can be a very good approach to evaluate force fields.

4. Ab Initio Protein Folding

As early as in 1975, Levitt & Warshel reported the folding simulation of a small protein (bovine pancreatic trypsin inhibitor) from fully extended conformation to RMSD of 5.3 Å, using a very simple model (two particles per residue) [113]. Limited by computational resource, energy minimization and normal-mode thermalisation were used instead of MD. Since then, there were numerous computational studies of folding and tertiary structure prediction of peptides and proteins, mostly using simplified protein models with Monte Carlo methods [114]. At early time, the direct atomistic MD simulation of folding has been impractical.

With increasing time resolution of experiments, people began to realize that some ultrafast folding proteins (with 20 ~ 106 residues) can fold on the time scale of a few μs [115]. This encourage the use MD simulation to study protein folding. In 1998, using a massively parallel supercomputer, a simulation of the folding of 36-residue villin headpiece subdomain (villin HP-36) in explicit water was performed for 1 μs [116]. Collapsed structures with native-like secondary structures were observed, and the RMSD of the representative structure is 5.7 Å. In 1999, Takada et al. reported successful folding of an artificial three-helix-bundle protein (54 residues) using overdamped Langevin dynamics (an MD version usually for implicit-solvent coarse-gained simulations) [117]. A simplified model of 3~4 particles per residue allowed trajectories up to 1 microsecond (μs). Soon later, a naturally occurring three-helix-bundle protein was folded from random-coil structures to RMSD of ~ 3 Å within 1 μs, by similar method [118].

Since 2000, successful atomistic folding simulations of some miniproteins in implicit solvent were reported, including: 20-residue three-stranded β-sheet Beta3s [119], 16-residue C-terminal β-hairpin from protein G [120], 23-residue BBA5 with ββα structure [121,122], 20-residue α-helical Trp-cage [123]. Besides, 35-residue villin HP [124], 10-residue chignolin [125], and 46-residue α-helical fragment B of protein A [126] were successfully folded using explicit solvent. Plain MD or REMD were used in these works, but other enhanced sampling methods can also be used. For example, accelerated MD has been used to fold four fast-folding proteins (chignolin, Trp-cage, villin HP, and WW domain) [127].

Besides, more efficient discrete molecular dynamics (DMD) simulation can be performed based on stepwise potentials with implicit solvation [128]. Using replica-exchange DMD, six small proteins (20–60 residues) with diverse native structures have been successfully folded [129]. For the smallest three (Trp-cage, villin HP, WW domain), predictions of RMSDs between 2~3 Å can be achieved. However, we do not see wide spread use of DMD in folding studies, possibly due to using less realistic physics model.

One approach to surmount the time scale barrier is to construct Markov state models (MSMs) using many different MD trajectories. It has been successfully used for all-atom *ab initio* folding of small systems such as the villin HP-35, for which the most populated state has an average RMSD of 2.3Å [130]. Combining thousands of MD simulations with explicit solvent (each trajectory up to 1 μs, totally 1.3 ms), an atomistic model of the folding of an 80-residue fragment of the λ repressor was obtained to capture dynamics on a 10 milliseconds time scale [131]. Using Folding@Home distributed computing platform [132], ~3000 unfolded-initiated trajectories of implicit-solvent MD were generated for 39-residue protein NTL-9 with an experimental folding time of ~1.5 ms, two trajectories reached RMSD < 3.5 Å [133]. An alternative method is Milestoning [134,135], developed by Elber and coworkers. Milestoning samples slow processes by coarse graining conformational space and performing large numbers of short simulations, yielding kinetics and pathways. However, these methods are too expensive to be applied in practical *ab initio* structure prediction, although very useful for folding mechanism studies.

With increasing computing power and force field improvements, simultaneously folding simulations of more proteins have been reported. Using special purpose computer *Anton*, Lindorff-Larsen et al. reported the first successful large-scale folding simulation [35]. Eleven proteins of 10~80 residues were reversibly folded to RMSD < 3.6 Å using the CHARMM22* force field in explicit water. Using REMD with our RSFF1 [136] force field, we also folded all these 11 proteins to RMSD < 3.8 Å, along with the Trp-cage TC5b and wild-type engrailed homeodomain (EnHD) [137] (Fig. 3). By analyzing continuous trajectories tracking every replica exchange, we also found that REMD can increase the folding rate by about 6 times, through significantly (> 10^2 times in most cases) increases the diffusion rate on rough energy landscape. Using inexpensive GPUs and implicit solvent model (~1 μs/day can be achieved), Simmerling and coworkers reported successful folding for 16 of 17 proteins (10–92 residues) with a variety of secondary structures and topologies, although the native conformations may not be thermodynamically preferred [138].

Although ultrafast folding proteins are relatively rare, a considerable fraction of protein domains can fold within time scale of milliseconds [139,140]. Using *Anton* machine, Piana et al. performed eight one-millisecond (1 ms) MD simulations of ubiquitin (a very common 76-residue protein) in explicit water [141]. Starting from the folded structure, spontaneously unfolding to RMSDs of > 20Å and refolding to Cα RMSDs of 0.5 Å was observed. However, no folding events were observed in the two simulations initiated in the unfolded state, which is understandable since the estimated folding time is about 3 ms.

From above, it is clear that atomistic MD simulation of folding can be used for *ab initio* PSP, but it is still quite expensive and do not have significant advantage compared with fragment-assembly and MC-based methods in real PSP application. Instead, it is often used to study the folding mechanism, which is also scientifically very important.

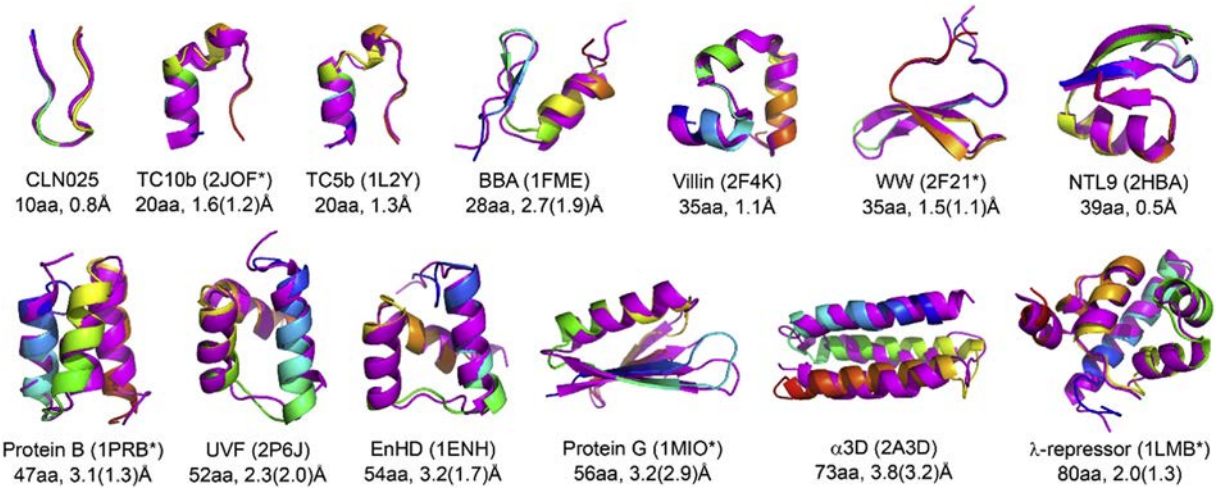

Fig. 3. Large-scale folding simulations reported in 2014 [137], using REMD with explicit solvent model. Experimental and predicted (center of the most populated cluster) structures are shown in magenta and rainbow color, respectively. PDB ID for each protein is given in parentheses with a star indicating a close variant was used. The number of amino acid (aa) residues and the Cα-RMSD (value in parenthesis is without a few terminal residues) of each predicted structure are also given.

However, when using coarse-gained force field (such as UNRES [142]), much more efficient conformational sampling can be achieved and larger protein can be folded, even in real *ab initio* PSP, rather than reproducing known structures. Especially, UNRES model with REMD simulation outperformed bioinformatics methods on predicting the correct topology of target T0663 (about 200 residues) in the CASP10 without any input from template or experiment [143,144].

5. Protein Structure Refinement

Currently, various informatics-based PSP methods have been developed. Usually, MD-based methods are not necessary, even when a detectable template is not available. However, currently predicted protein models are often much less accurate compared with experiments, limiting their usage in some important applications including structure-based drug design [18,145]. Thus, structure refinement from low-accuracy models to high-accurate ones is very important, although it is also a very difficult challenge [146]. A crucial motivation for developing physics-based refinement methods is that if the limits of informatics-based methods were being reached, utilizing physics was essential to finally have modeling methods that rival experiment.

In 2000, refinement using explicit-solvent MD simulations that employ the locally enhanced sampling (LES) was applied to low-resolution model of a small disulfide-rich 29-residue protein CMTI-1, and improvement from 3.7 Å to 2.5 Å was reported [147]. Explicit-solvent MD simulation of Rosetta models (of 36-mer HP-36 and 65-mer S15) was also used to generate structures for subsequent ranking using the MM-PBSA free energy function [148], and structures with RMSD < 1.5 Å can be sampled. However, a systematic study (in 2001) on 12 small, single-domain proteins failed to observe successful refinement [149]. On a set of 20 proteins, explicit-solvent MD simulations of most proteins moved initial model structures further away from their native conformations, with performances worse than energy minimization using implicit-solvent potentials [150]. Several studies reveal the importance of proper scoring functions to select more native-like ones among structures sampled during MD [151–153], because the percentage of more native-like (improved) structures is usually <50% and may decrease as refinement progress.

Restraints can be applied to focus the conformational sampling on the vicinity of the initial model. In 2007, Chen & Brooks performed implicit-solvent REMD simulations with distance restraints between Cα atoms on five protein models, and significantly refinements were observed on three of them [154]. Later, Feig & coworkers [155,156] made significant progress, using restrained explicit-solvent MD and special structure selection and averaging protocol assisted by a statistical potential (DFIRE [157]). Their method ranked the first in the model refinement category of CASP10 [146]. Lee's group also developed a refinement protocol based on a series of short (5 ns in total) explicit-solvent MD simulations with weak positional and distance restraints [158]. Recently, the combination of restrained MD simulations with accurate force fields is clearly useful and has been adopted by most top-ranking groups in the CASP12 refinement challenge [159]. However, the top-ranking groups are relatively conservative, yielding structures that are quite close to the initial ones.

Shaw and colleagues found that their CHARMM22* force field, which can fold a diverse set of small proteins, may not stabilize the experimental structure of a protein in long-time MD simulations, and good refinement can hardly be achieved [160]. Thus, the success of structure refinement highly relies on the force field accuracy. We recently evaluated the applicability of RSFF1 in protein structure refinement [161]. For 30 single-domain proteins from CASP8-10 refinement targets with diverse structures and a large C_α RMSD coverage of 1-9 Å, MD simulations (380 K) with weak C_α position restraints gave best structures with RMSD reduced by -0.85 Å on average. Using long-time REMD simulations with RSFF1, two homology models (TR614 and TR624) with initial RMSD > 5 Å, can be improved to RMSD < 3 Å. Results from CASP12 indicate that our approach is adventurous, and can provide significantly refined models for some targets (such as TR866, TR894 and TR944) but performs modestly overall.

6. Data-assisted Modeling

As described above, applications of MD in both *ab initio* structure prediction and model refinement [162] suffer from two interrelated challenges: insufficient conformational sampling and inaccurate force field. However, limited amounts of structural information can accelerate MD-based structure determination and may also improve simulation accuracy [163]. Meanwhile, experimental techniques have been developed to provide limited (low-resolution, sparse, ambiguous, or uncertain) structure information in a relatively short time and low cost, and bioinformatics techniques have also been developed to predict structure information including secondary structures and residue-residue contacts. Data-assisted modeling has become a sub-category of the CASP experiments since CASP11 (2014) [164].

It has been a long history to use data from known protein structures to guide MD folding simulation and structure prediction. In 1989, Friedrichs and Wolynes proposed the "associative memory Hamiltonians" (AMH) [165], which can learn structure features from a set of

memory proteins. With the incorporation of homologous protein(s) in the memory set and certain information based on secondary structure prediction, near-native structures of a 111-residue protein (rice cytochrome *c*) can be obtained starting from random structures using simulated annealing MD with a simplified model [166]. With AMH constructed from a database of non-homologous proteins, several α-helical proteins [167] and α/β proteins [168] can be folded to near-native structures (4–8 Å) using MD-based simulated annealing. Further improvements were made by the same group, especially the incorporation of water-mediated interactions [169]. After years of developments, a coarse-grained protein force field called AWSEM (associative memory, water mediated, structure and energy model) was established, which incorporates local structure biasing from fragments with known structures and similar sequences [170]. However, it seems not to be superior to popular fragment-assembly-based *de novo* PSP methods. Recently, a new scheme called AAWSEM (atomistic associative memory, water mediated, structure and energy model) has been developed [171]. It is an *ab initio* PSP method that starts from the ground up without using bioinformatics input.

MD-based methods have also been developed to incorporate distance distribution derived from the SAXS experiment as restraints, including the use of coarse-grained force field [172]. Restraints from various paramagnetic NMR experiments (including pseudo-contact shifts, residual dipolar couplings, paramagnetic relaxation enhancement, and cross-correlated relaxation) can be incorporated in computer modeling of protein 3D atomic structures, but substantial challenges remain before wide spread use [173].

Chemical crosslinking mass spectrometry (XL-MS) can provide information about residue-pairs in close proximity that can be incorporated into modeling, although the data may be spares and of low-resolution. Replica-exchange DMD simulations using the Medusa potential with distance restraints from XL-MS experiments gave lowest-energy models of 2.7 Å and 5.4 Å for FK506-binding protein and myoglobin, respectively [174]. MD simulations of a large number of proteins can also be used to find appropriate distance constraint from investigating Lys side-chain motions [175].

Dill and coworkers developed a method called Modeling Employing Limited Data (MELD) that can harness problematic experimental or theoretical data in a Bayesian framework to assist physics-based structure modeling [176,177], which can use a variety of sampling methods obeying detailed balance but implicit-solvent REMD is a good choice. Using loose physical insights (such as proteins have hydrophobic cores and secondary structures), MD simulations of protein folding can be speed up by two ~ five orders of magnitude [178]. Therefore, the MELD method can also be used for *ab initio* PSP, and structures of three proteins (with 97, 67, 68 residues, respectively) from CASP targets can be predicted blindly with RMSD of 2.8 Å , 1.4 Å, 1.5 Å from native structures, respectively [179].

Besides altering the potential energy function in MD, an iterative screening-after-sampling strategy can be used [180]. By selecting conformations that better fit with the low-resolution data from each cycle of MD simulations, high-quality atomic model can be achieved.

7. Summary

With rapid increasing of computer performance, as well as continuous software and force field developments, MD simulation has been increasingly used in studying biomolecular systems. In principle, it can describe the underline physics of detailed atomic interactions determining a protein structure, and potentially be more accurate than knowledge-based PSP methods. In this mini-review, we showed that all-atom MD simulations can predict structures of cyclic peptides and other peptide-based foldamers with accuracy similar to experiments. Then, we summarized some notable successes in reproducing experimental 3D structures of small proteins through *ab initio* folding simulations. We also described recent advancements of using MD simulations with state-of-the-art force fields in improving structure models from bioinformatics-based PSP, which is one of the most useful for real-word applications. Finally, some methodology developments and applications of using limited experimental or theoretical data to guide MD-based structure modeling were also introduced. We feel that, in the future, more sophisticated and integrative methods will be developed, including those combining different levels of structure representation (multi-scale MD simulations) and those utilizing the power of machine learning (to take advantage of large amount of data generated by MD) [181].

Declaration of Competing Interest

The authors claim no conflict of interest.

Acknowledgements

We thank the financial supports from the National Natural Science Foundation of China (21573009), and the Shenzhen Science and Technology Innovation Committee (JCYJ20170412150507046, JCYJ20170412151002616).

References

[1] Westbrook JD, Burley SK. How structural biologists and the protein data bank contributed to recent FDA new drug approvals. Structure 2019;27:211–7. https://doi.org/10.1016/j.str.2018.11.007.
[2] Kendrew JC, Bodo G, Dintzis HM, Parrish RG, Wyckoff H, Phillips DC. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. Nature 1958;181:662–6. https://doi.org/10.1038/181662a0.
[3] Kendrew JC, Dickerson RE, Strandberg BE, Hart RG, Davies DR, Phillips DC, et al. Structure of Myoglobin: a three-dimensional fourier synthesis at 2 Å. Resol Nat 1960;185:422–7. https://doi.org/10.1038/185422a0.
[4] Wüthrich K. The way to NMR structures of proteins. Nat Struct Biol 2001;8:923–5. https://doi.org/10.1038/nsb1101-923.
[5] Carpenter EP, Beis K, Cameron AD, Iwata S. Overcoming the challenges of membrane protein crystallography. Curr Opin Struct Biol 2008;18:581–6. https://doi.org/10.1016/j.sbi.2008.07.001.
[6] Shi Y. A glimpse of structural biology through X-ray crystallography. Cell 2014;159:995–1014. https://doi.org/10.1016/j.cell.2014.10.051.
[7] Miao J, Ishikawa T, Robinson IK, Murnane MM. Suplementary Materials for Beyond crystallography: diffractive imaging using coherent x-ray light sources. Science 2015;348:530–5. https://doi.org/10.1126/science.aaa1394.
[8] Chen Bai X, McMullan G, SHW Scheres. How cryo-EM is revolutionizing structural biology. Trends Biochem Sci 2015;40:49–57. https://doi.org/10.1016/j.tibs.2014.10.005.
[9] Cheng Y. Membrane protein structural biology in the era of single particle cryo-EM. Curr Opin Struct Biol 2018;52:58–63. https://doi.org/10.1016/j.sbi.2018.08.008.
[10] Reuter JA, Spacek DV, Snyder MP. High-throughput sequencing technologies. Mol Cell 2015;58:586–97. https://doi.org/10.1016/j.molcel.2015.05.004.
[11] Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet 2016;17:333–51. https://doi.org/10.1038/nrg.2016.49.
[12] John B, Sali A. Comparative protein structure modeling by iterative alignment, model building and model assessment. Nucleic Acids Res 2003;31:3982–92. https://doi.org/10.1093/nar/gkg460.
[13] Bradley P, Misura KMS, Baker D. Toward high-resolution de novo structure prediction for small proteins. Science 2005;309:1868–71. https://doi.org/10.1126/science.1113801.
[14] Rohl CA, Strauss CEM, Misura KMS, Baker D. Protein structure prediction using Rosetta. Methods Enzymol 2004;383:66–93. https://doi.org/10.1016/S0076-6879(04)83004-0.
[15] Xu D, Zhang Y. Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field. Proteins 2012;80:1715–35. https://doi.org/10.1002/prot.24065.
[16] Marks DS, Colwell LJ, Sheridan R, Hopf TA, Pagnani A, Zecchina R, et al. Protein 3D structure computed from evolutionary sequence variation. PLoS One 2011;6:e28766. https://doi.org/10.1371/journal.pone.0028766.
[17] Ovchinnikov S, Park H, Varghese N, Huang P-S, Pavlopoulos GA, Kim DE, et al. Protein structure determination using metagenome sequence data. Science 2017;355:294–8. https://doi.org/10.1126/science.aah4043.
[18] Zhang Y. Protein structure prediction: when is it useful? Curr Opin Struct Biol 2009;19:145–55. https://doi.org/10.1016/j.sbi.2009.02.005.
[19] Epstein CJ, Goldberger RF, Anfinsen CB. The genetic control of tertiary protein structure: studies with model systems. Cold Spring Harb Symp Quant Biol 2012;28:439–49. https://doi.org/10.1101/sqb.1963.028.01.060.

[20] Anfinsen CB. Principles that govern the folding of protein chains. Science 1973;181:223–30. https://doi.org/10.1126/science.181.4096.223.

[21] Ulmschneider JP, Jorgensen WL. Polypeptide folding using monte carlo sampling, concerted rotation, and continuum solvation. J Am Chem Soc 2004;126:1849–57. https://doi.org/10.1021/ja0378862.

[22] Meinke JH, Hansmann UHE. Free-energy-driven folding and thermodynamics of the 67-residue protein GS-α 3 W-A large-scale Monte Carlo study. J Comput Chem 2009;30:1642–8. https://doi.org/10.1002/jcc.21321.

[23] Xu J, Huang L, Shakhnovich EI. The ensemble folding kinetics of the FBP28 WW domain revealed by an all-atom Monte Carlo simulation in a knowledge-based potential. Proteins Struct Funct Bioinforma 2011;79:1704–14. https://doi.org/10.1002/prot.22993.

[24] Dill KA, Bromberg S, Yue K, Chan HS, Ftebig KM, Yee DP, et al. Principles of protein folding - a perspective from simple exact models. Protein Sci 2008;4:561–602. https://doi.org/10.1002/pro.5560040401.

[25] Derreumaux P. From polypeptide sequences to structures using Monte Carlo simulations and an optimized potential. J Chem Phys 1999;111:2301–10. https://doi.org/10.1063/1.479501.

[26] Kmiecik S, Kouza M, Badaczewska-Dawid A, Kloczkowski A, Kolinski A. Modeling of Protein Structural Flexibility and Large-Scale Dynamics: coarse-Grained Simulations and Elastic Network Models. Int J Mol Sci 2018;19:3496. https://doi.org/10.3390/ijms19113496.

[27] Kmiecik S, Kolinski A. Characterization of protein-folding pathways by reduced-space modeling. Proc Natl Acad Sci 2007;104:12330–5. https://doi.org/10.1073/pnas.0702265104.

[28] Kmiecik S, Kolinski A. Folding pathway of the B1 Domain of Protein G explored by multiscale modeling. Biophys J 2008;94:726–36. https://doi.org/10.1529/biophysj.107.116095.

[29] Cui M, Mezei M, Osman R. Prediction of protein loop structures using a local move Monte Carlo approach and a grid-based force field. Protein Eng Des Sel 2008;21:729–35. https://doi.org/10.1093/protein/gzn056.

[30] Alder BJ, Wainwright TE. Studies in molecular dynamics I. General method. J Chem Phys 1959;31:459–66. https://doi.org/10.1063/1.1730376.

[31] Alder BJ, Wright TEWA. Studies in molecular dynamics. II. Behavior of a small number of elastic spheres. J Chem Phys 1960;33:1439–51. https://doi.org/10.1063/1.1731425.

[32] Rahman A. Correlations in the motion of atoms in Liquid Argon. Phys Rev 1964;136:A405–11. https://doi.org/10.1103/PhysRev.136.A405.

[33] Vendruscolo M, Dobson CM. Protein dynamics: moore's Law in molecular biology. Curr Biol 2011;21:R68–70. https://doi.org/10.1016/j.cub.2010.11.062.

[34] Perez A, Morrone JA, Simmerling C, Dill KA. Advances in free-energy-based simulations of protein folding and ligand binding. Curr Opin Struct Biol 2016;36:25–31. https://doi.org/10.1016/j.sbi.2015.12.002.

[35] Lindorff-Larsen K, Piana S, Dror RO, Shaw DE. How fast-folding proteins fold. Science 2011;334:517–20. https://doi.org/10.1126/science.1208351.

[36] McCammon JA, Gelin BR, Karplus M. Dynamics of folded proteins. Nature 1977;267:585–90. https://doi.org/10.1038/267585a0.

[37] Gaalswyk K, Muniyat MI, MacCallum JL. The emerging role of physical modeling in the future of structure determination. Curr Opin Struct Biol 2018;49:145–53. https://doi.org/10.1016/j.sbi.2018.03.005.

[38] Chouard T. Structural biology: Breaking the protein rules. Nature 2011;471:151–3. https://doi.org/10.1038/471151a.

[39] Best RB. Computational and theoretical advances in studies of intrinsically disordered proteins. Curr Opin Struct Biol 2017;42:147–54. https://doi.org/10.1016/j.sbi.2017.01.006.

[40] Allen F, Almasi G, Andreoni W, Beece D, Berne BJ, Bright A, et al. Blue gene: a vision for protein science using a petaflop supercomputer. IBM Syst J 2001;40:310–27. https://doi.org/10.1147/sj.402.0310.

[41] Hess B, Kutzner C, van der Spoel D, Lindahl E. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. J Chem Theory Comput 2008;4:435–47. https://doi.org/10.1021/ct700301q.

[42] Freddolino PL, Arkhipov AS, Larson SB, McPherson A, Schulten K. Molecular dynamics simulations of the complete Satellite Tobacco Mosaic Virus. Structure 2006;14:437–49. https://doi.org/10.1016/j.str.2005.11.014.

[43] Shaw DE, Chao JC, Eastwood MP, Gagliardo J, Grossman JP, Ho CR, et al. Anton, a special-purpose machine for molecular dynamics simulation. Commun ACM 2008;51:91. https://doi.org/10.1145/1364782.1364802.

[44] Shaw DE, Maragakis P, Lindorff-Larsen K, Piana S, Dror RO, Eastwood MP, et al. Atomic-level characterization of the structural dynamics of Proteins. Science 2010;330:341–6. https://doi.org/10.1126/science.1187409.

[45] Liu W, Schmidt B, Voss G, Müller-Wittig W. Accelerating molecular dynamics simulations using graphics processing units with CUDA. Comput Phys Commun 2008;179:634–41. https://doi.org/10.1016/j.cpc.2008.05.008.

[46] Harvey MJ, De Fabritiis G. A survey of computational molecular science using graphics processing units. Wiley Interdiscip Rev Comput Mol Sci 2012;2:734–42. https://doi.org/10.1002/wcms.1101.

[47] Salomon-Ferrer R, Götz AW, Poole D, Le Grand S, Walker RC. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle Mesh Ewald. J Chem Theory Comput 2013;9:3878–88. https://doi.org/10.1021/ct400314y.

[48] Freddolino PL, Liu F, Gruebele M, Schulten K. Ten-microsecond molecular dynamics simulation of a fast-folding WW domain. Biophys J 2008;94:L75–7. https://doi.org/10.1529/biophysj.108.131565.

[49] Freddolino PL, Park S, Roux B, Schulten K. Force field bias in protein folding simulations. Biophys J 2009;96:3772–80. https://doi.org/10.1016/j.bpj.2009.02.033.

[50] Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, et al. A second generation force field for the simulation of proteins, nucleic acids, and Organic molecules. J Am Chem Soc . 1995 , 117 , 5179—5197 J Am Chem Soc 1996;118. https://doi.org/10.1021/ja955032e 2309–2309.

[51] Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins 2006;65:712–25. https://doi.org/10.1002/prot.21123.

[52] Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, et al. Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins 2010;78:1950–8. https://doi.org/10.1002/prot.22711.

[53] Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: improving the accuracy of Protein Side chain and backbone parameters from ff99SB. J Chem Theory Comput 2015;11:3696–713. https://doi.org/10.1021/acs.jctc.5b00255.

[54] MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 1998;102:3586–616. https://doi.org/10.1021/jp973084f.

[55] Mackerell AD, Feig M, Brooks CL. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. J Comput Chem 2004;25:1400–15. https://doi.org/10.1002/jcc.20065.

[56] Piana S, Lindorff-Larsen K, Shaw DE. How robust are protein folding simulations with respect to force field parameterization? Biophys J 2011;100:L47–9. https://doi.org/10.1016/j.bpj.2011.03.051.

[57] Best RB, Zhu X, Shim J, Lopes PEM, Mittal J, Feig M, et al. Optimization of the additive CHARMM All-Atom protein force field targeting improved sampling of the backbone ϕ, ψ and Side-Chain χ 1 and χ 2 Dihedral Angles. J Chem Theory Comput 2012;8:3257–73. https://doi.org/10.1021/ct300400x.

[58] Jorgensen WL, Maxwell DS, Tirado-Rives J. Development and testing of the OPLS All-Atom force field on conformational energetics and properties of organic liquids. J Am Chem Soc 1996;118:11225–36. https://doi.org/10.1021/ja9621760.

[59] Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with Accurate Quantum chemical calculations on peptides †. J Phys Chem B 2001;105:6474–87. https://doi.org/10.1021/jp003919d.

[60] Robertson MJ, Tirado-Rives J, Jorgensen WL. Improved Peptide and Protein torsional energetics with the OPLS-AA force field. J Chem Theory Comput 2015;11:3499–509. https://doi.org/10.1021/acs.jctc.5b00356.

[61] Li H, Tang C, Wingreen NS. Nature of driving force for protein folding: a result from analyzing the statistical potential. Phys Rev Lett 1997;79:765–8. https://doi.org/10.1103/PhysRevLett.79.765.

[62] Bellissent-Funel M-C, Hassanali A, Havenith M, Henchman R, Pohl P, Sterpone F, et al. Water determines the structure and dynamics of proteins. Chem Rev 2016;116:7673–97. https://doi.org/10.1021/acs.chemrev.5b00664.

[63] Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys 1983;79:926–35. https://doi.org/10.1063/1.445869.

[64] Still WC, Tempczyk A, Hawley RC, Hendrickson T. Semianalytical treatment of solvation for molecular mechanics and dynamics. J Am Chem Soc 1990;112:6127–9. https://doi.org/10.1021/ja00172a038.

[65] Onufriev A, Bashford D, Case DA. Exploring protein native states and large-scale conformational changes with a modified generalized born model. Proteins 2004;55:383–94. https://doi.org/10.1002/prot.20033.

[66] Nguyen H, Roe DR, Simmerling C. Improved generalized born solvent model parameters for protein simulations. J Chem Theory Comput 2013;9:2020–34. https://doi.org/10.1021/ct3010485.

[67] Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA. The weighted histogram analysis method for free-energy calculations on biomolecules I the method. J Comput Chem 1992;13:1011–21. https://doi.org/10.1002/jcc.540130812.

[68] Laio A, Parrinello M. Escaping free-energy minima. Proc Natl Acad Sci 2002;99:12562–6. https://doi.org/10.1073/pnas.202427399.

[69] Hamelberg D, Mongan J, McCammon JA. Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. J Chem Phys 2004;120:11919–29. https://doi.org/10.1063/1.1755656.

[70] Sugita Y, Okamoto Y. Replica-exchange molecular dynamics method for protein folding. Chem Phys Lett 1999;314:141–51. https://doi.org/10.1016/S0009-2614(99)01123-9.

[71] Gao YQ. An integrate-over-temperature approach for enhanced sampling. J Chem Phys 2008;128:64105. https://doi.org/10.1063/1.2825614.

[72] Wu S, Skolnick J, Zhang Y. Ab initio modeling of small proteins by iterative TASSER simulations. BMC Biol 2007;5:17. https://doi.org/10.1186/1741-7007-5-17.

[73] Henninot A, Collins JC, Nuss JM. The current state of Peptide drug discovery: back to the future? J Med Chem 2017. https://doi.org/10.1021/acs.jmedchem.7b00318 acs.jmedchem.7b00318.

[74] Flaherty RA, Freed SD, Lee SW. The wide world of ribosomally encoded Bacterial Peptides. PLoS Pathog 2014;10:e1004221. https://doi.org/10.1371/journal.ppat.1004221.

[75] Gould A, Camarero JA. Cyclotides: overview and biotechnological applications. Chem Bio Chem 2017;18:1350–63. https://doi.org/10.1002/cbic.201700153.

[76] Akondi KB, Muttenthaler M, Dutertre S, Kaas Q, Craik DJ, Lewis RJ, et al. Discovery, synthesis, and structure–activity relationships of Conotoxins. Chem Rev 2014;114:5815–47. https://doi.org/10.1021/cr400401e.

[77] Gao B, Peng C, Yang J, Yi Y, Zhang J, Shi Q. Cone snails: a big store of Conotoxins for novel drug discovery. Toxins (Basel) 2017;9:397. https://doi.org/10.3390/toxins9120397.

[78] Geng L, Wang Z, Yang X, Li D, Lian W, Xiang Z, et al. Structure-based design of Peptides with high affinity and specificity to HER2 Positive Tumors. Theranostics 2015; 5:1154–65. https://doi.org/10.7150/thno.12398.

[79] Pelay-Gimeno M, Glas A, Koch O, Grossmann TN. Structure-based design of inhibitors of Protein-Protein Interactions: mimicking Peptide binding Epitopes. Angew Chem Int Ed 2015;54:8896–927. https://doi.org/10.1002/anie.201412070.

[80] Hu X, Nguyen KT, Verlinde CLMJ, Hol WGJ, Pei D. Structure-based design of a Macrocyclic Inhibitor for Peptide Deformylase. J Med Chem 2003;46:3771–4. https://doi.org/10.1021/jm034113f.

[81] Wang CK, Craik DJ. Designing macrocyclic disulfide-rich peptides for biotechnological applications. Nat Chem Biol 2018;14:417–27. https://doi.org/10.1038/s41589-018-0039-y.

[82] Gurrath M, Muller G, Kessler H, Aumailley M, Timpl R. Conformation/activity studies of rationally designed potent anti-adhesive RGD peptides. Eur J Biochem 1992; 210:911–21. https://doi.org/10.1111/j.1432-1033.1992.tb17495.x.

[83] Bock JE, Gavenonis J, Kritzer JA. Getting in shape: controlling peptide bioactivity and bioavailability using conformational constraints. ACS Chem Biol 2013;8: 488–99. https://doi.org/10.1021/cb300515u.

[84] Das R. Four small puzzles that Rosetta doesn't solve. PLoS One 2011;6:e20044. https://doi.org/10.1371/journal.pone.0020044.

[85] Beaufays J, Lins L, Thomas A, Brasseur R. In silico predictions of 3D structures of linear and cyclic peptides with natural and non-proteinogenic residues. J Pept Sci 2012;18:17–24. https://doi.org/10.1002/psc.1410.

[86] Merten C, Li F, Bravo-Rodriguez K, Sanchez-Garcia E, Xu Y, Sander W. Solvent-induced conformational changes in cyclic peptides: a vibrational circular dichroism study. Phys Chem Chem Phys 2014;16:5627–33. https://doi.org/10.1039/C3CP55018D.

[87] Cuniasse P, Raynal I, Yiotakis A, Dive V. Accounting for conformational variability in NMR Structure of Cyclopeptides: ensemble averaging of interproton distance and coupling constant restraints. J Am Chem Soc 1997;119:5239–48. https://doi.org/10.1021/ja9636810.

[88] McHugh SM, Rogers JR, Yu H, Lin YS. Insights into how Cyclic Peptides switch conformations. J Chem Theory Comput 2016;12:2480–8. https://doi.org/10.1021/acs.jctc.6b00193.

[89] Shen Y, Maupetit J, Derreumaux P, Tufféry P. Improved PEP-FOLD approach for Peptide and Miniprotein structure prediction. J Chem Theory Comput 2014;10: 4745–58. https://doi.org/10.1021/ct500592m.

[90] McHugh SM, Yu H, Slough DP, Lin Y-S. Mapping the sequence–structure relationships of simple cyclic hexapeptides. Phys Chem Chem Phys 2017;19:3315–24. https://doi.org/10.1039/C6CP06192C.

[91] Ho BK, Dill KA. Folding very short Peptides using molecular dynamics. PLoS Comput Biol 2006;2:e27. https://doi.org/10.1371/journal.pcbi.0020027.

[92] Voelz VA, Shell MS, Dill KA. Predicting Peptide structures in native proteins from physical simulations of fragments. PLoS Comput Biol 2009;5:e1000281. https://doi.org/10.1371/journal.pcbi.1000281.

[93] Daura X, Jaun B, Seebach D, van Gunsteren WF, Mark AE. Reversible peptide folding in solution by molecular dynamics simulation. J Mol Biol 1998;280:925–32. https://doi.org/10.1006/jmbi.1998.1885.

[94] Daura X, Gademann K, Jaun B, Seebach D, Van Gunsteren WF, Mark AE. Peptide folding: when simulation meets experiment. Angew Chem Int Ed 1999;38: 236–40. https://doi.org/10.1002/(SICI)1521-3773(19990115)38:1/2<236::AID-ANIE236>3.0.CO;2-M.

[95] Damas JM, Filipe LCS, Campos SRR, Lousa D, Victor BL, Baptista AM, et al. Predicting the thermodynamics and kinetics of Helix Formation in a Cyclic Peptide model. J Chem Theory Comput 2013;9:5148–57. https://doi.org/10.1021/ct400529k.

[96] Razavi AM, Wuest WM, Voelz VA. Computational screening and selection of cyclic peptide hairpin mimetics by molecular simulation and kinetic network models. J Chem Inf Model 2014;54:1425–32. https://doi.org/10.1021/ci500102y.

[97] Wakefield AE, Wuest WM, Voelz VA. Molecular simulation of conformational preorganization in cyclic RGD peptides. J Chem Inf Model 2015;55:806–13. https://doi.org/10.1021/ci500768u.

[98] Butterfoss GL, Yoo B, Jaworski JN, Chorny I, Dill KA, Zuckermann RN, et al. De novo structure prediction and experimental characterization of folded peptoid oligomers. Proc Natl Acad Sci 2012;109:14320–5. https://doi.org/10.1073/pnas.1209945109.

[99] Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and testing of a general amber force field. J Comput Chem 2004;25:1157–74. https://doi.org/10.1002/jcc.20035.

[100] Yedvabny E, Nerenberg PS, So C, Head-Gordon T. Disordered structural ensembles of Vasopressin and Oxytocin and their mutants. J Phys Chem B 2015;119:896–905. https://doi.org/10.1021/jp505902m.

[101] Yu H, Lin Y-S. Toward structure prediction of cyclic peptides. Phys Chem Chem Phys 2015;17:4210–9. https://doi.org/10.1039/C4CP04580G.

[102] Geng H, Jiang F, Wu Y-D. Accurate structure prediction and conformational analysis of cyclic Peptides with Residue-Specific force fields. J Phys Chem Lett 2016. https://doi.org/10.1021/acs.jpclett.6b00452 acs.jpclett.6b00452.

[103] Jiang F, Geng H. Computational methods for studying conformational behaviors of cyclic Peptides. Methods Mol Biol 2001;2019:61–71. https://doi.org/10.1007/978-1-4939-9504-2_4.

[104] Zhou C-Y, Jiang F, Wu Y-D. Residue-specific force field based on Protein Coil library. RSFF2: modification of AMBER ff99SB. J Phys Chem B 2015;119:1035–47. https://doi.org/10.1021/jp5064676.

[105] Slough DP, McHugh SM, Cummings AE, Dai P, Pentelute BL, Kritzer JA, et al. Designing well-structured cyclic Pentapeptides based on sequence-structure relationships. J Phys Chem B 2018;122:3908–19. https://doi.org/10.1021/acs.jpcb.8b01747.

[106] Hu K, Geng H, Zhang Q, Liu Q, Xie M, Sun C, et al. An in-tether chiral center modulates the helicity, cell permeability, and target binding affinity of a Peptide. Angew Chem Int Ed 2016;55:8013–7. https://doi.org/10.1002/anie.201602806.

[107] Zhao H, Liu Q-S, Geng H, Tian Y, Cheng M, Jiang Y-H, et al. Crosslinked Aspartic acids as Helix-Nucleating templates. Angew Chem Int Ed 2016;55:12088–93. https://doi.org/10.1002/anie.201606833.

[108] Hou Z, Sun C, Geng H, Hu K, Xie M, Ma Y, et al. Facile chemoselective modification of Thio-Ethers generates Chiral center-induced Helical Peptides. Bioconjug Chem 2018;29:2904–8. https://doi.org/10.1021/acs.bioconjchem.8b00624.

[109] Slough DP, Yu H, McHugh SM, Lin Y-S. Toward accurately modeling N-methylated cyclic peptides. Phys Chem Chem Phys 2017;19:5377–88. https://doi.org/10.1039/C6CP07700E.

[110] Izmailov SA, Podkorytov IS, Skrynnikov NR. Simple MD-based model for oxidative folding of peptides and proteins. Sci Rep 2017;7:9293. https://doi.org/10.1038/s41598-017-09229-7.

[111] Paissoni C, Nardelli F, Zanella S, Curnis F, Belvisi L, Musco G, et al. Critical assessment of force fields accuracy against NMR data for cyclic peptides containing β-amino acids. Phys Chem Chem Phys 2018:15807–16. https://doi.org/10.1039/C8CP00234G.

[112] Kamenik AS, Lessel U, Fuchs JE, Fox T, Liedl KR. Peptidic macrocycles - conformational sampling and thermodynamic characterization. J Chem Inf Model 2018;58: 982–92. https://doi.org/10.1021/acs.jcim.8b00097.

[113] Levitt M, Warshel A. Computer simulation of protein folding. Nature 1975;253: 694–8. https://doi.org/10.1038/253694a0.

[114] Skolnick J, Kolinski A. Computer simulations of globular Protein folding and tertiary structure. Annu Rev Phys Chem 2003;40:207–35. https://doi.org/10.1146/annurev.pc.40.100189.001231.

[115] Kubelka J, Hofrichter J, Eaton WA. The protein folding "speed limit". Curr Opin Struct Biol 2004;14:76–88. https://doi.org/10.1016/j.sbi.2004.01.013.

[116] Duan Y, Kollman PA. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. Science 1998;282:740–4. https://doi.org/10.1126/science.282.5389.740.

[117] Takada S, Luthey-Schulten Z, Wolynes PG. Folding dynamics with nonadditive forces: a simulation study of a designed helical protein and a random heteropolymer. J Chem Phys 1999;110:11616–29. https://doi.org/10.1063/1.479101.

[118] Takada S. Protein folding simulation with solvent-induced force field: folding pathway ensemble of three-helix-bundle proteins. Proteins Struct Funct Genet 2001; 42:85–98. https://doi.org/10.1002/1097-0134(20010101)42:1<85::AID-PROT90>3.0.CO;2-3.

[119] Ferrara P, Caflisch A. Folding simulations of a three-stranded antiparallel beta -sheet peptide. Proc Natl Acad Sci 2000;97:10780–5. https://doi.org/10.1073/pnas.190324897.

[120] Zagrovic B, Sorin EJ, Pande V. β-hairpin folding simulations in atomistic detail using an implicit solvent model 1 1Edited by F. Cohen. J Mol Biol 2001;313:151–69. https://doi.org/10.1006/jmbi.2001.5033.

[121] Snow CD, Nguyen H, Pande VS, Gruebele M. Absolute comparison of simulated and experimental protein-folding dynamics. Nature 2002;420:102–6. https://doi.org/10.1038/nature01160.

[122] Rhee YM, Pande VS. Multiplexed-replica exchange molecular dynamics method for Protein folding simulation. Biophys J 2003;84:775–86. https://doi.org/10.1016/S0006-3495(03)74897-8.

[123] Pitera JW, Swope W. Understanding folding and design: replica-exchange simulations of "Trp-cage" miniproteins. Proc Natl Acad Sci 2003;100:7587–92. https://doi.org/10.1073/pnas.1330954100.

[124] Freddolino PL, Schulten K. Common structural transitions in Explicit-solvent simulations of Villin Headpiece folding. Biophys J 2009;97:2338–47. https://doi.org/10.1016/j.bpj.2009.08.012.

[125] Seibert MM, Patriksson A, Hess B, van der Spoel D. Reproducible Polypeptide folding and structure prediction using molecular dynamics simulations. J Mol Biol 2005;354:173–83. https://doi.org/10.1016/j.jmb.2005.09.030.

[126] Garcia AE, Onuchic JN. Folding a protein in a computer: an atomic description of the folding/unfolding of Protein A. Proc Natl Acad Sci 2003;100:13898–903. https://doi.org/10.1073/pnas.2335541100.

[127] Miao Y, Feixas F, Eun C, McCammon JA. Accelerated molecular dynamics simulations of protein folding. J Comput Chem 2015;36:1536–49. https://doi.org/10.1002/jcc.23964.

[128] Proctor EA, Ding F, Dokholyan NV. Discrete molecular dynamics. Wiley Interdiscip Rev Comput Mol Sci 2011;1:80–92. https://doi.org/10.1002/wcms.4.

[129] Ding F, Tsao D, Nie H, Dokholyan NV. Ab Initio folding of Proteins with All-Atom discrete molecular dynamics. Structure 2008;16:1010–8. https://doi.org/10.1016/j.str.2008.03.013.

[130] Bowman GR, Beauchamp KA, Boxer G, Pande VS. Progress and challenges in the automated construction of Markov state models for full protein systems. J Chem Phys 2009;131. https://doi.org/10.1063/1.3216567.

[131] Bowman GR, Voelz VA, Pande VS. Atomistic folding simulations of the Five-Helix Bundle Protein λ 6−85. J Am Chem Soc 2011;133:664–7. https://doi.org/10.1021/ja106936n.

[132] Shirts M, Pande VS. COMPUTING: screen savers of the World Unite! Science 2000; 290:1903–4. https://doi.org/10.1126/science.290.5498.1903.

[133] Voelz VA, Bowman GR, Beauchamp K, Pande VS. Molecular simulation of Ab Initio Protein folding for a Millisecond Folder NTL9(1−39). J Am Chem Soc 2010;132: 1526–8. https://doi.org/10.1021/ja9090353.

[134] Faradjian AK, Elber R. Computing time scales from reaction coordinates by milestoning. J Chem Phys 2004;120:10880–9. https://doi.org/10.1063/1.1738640.

[135] Elber R. Perspective: computer simulations of long time dynamics. J Chem Phys 2016;144:60901. https://doi.org/10.1063/1.4940794.
[136] Jiang F, Zhou C-Y, Wu Y. Residue-specific force field based on the Protein Coil Library. RSFF1: modification of OPLS-AA/L. J Phys Chem B 2014;118:6983–98. https://doi.org/10.1021/jp5017449.
[137] Jiang F, Wu YD. Folding of fourteen small proteins with a residue-specific force field and replica-exchange molecular dynamics. J Am Chem Soc 2014;136:9536–9. https://doi.org/10.1021/ja502735c.
[138] Nguyen H, Maier J, Huang H, Perrone V, Simmerling C. Folding simulations for Proteins with diverse topologies are accessible in days with a Physics-Based Force Field and Implicit Solvent. J Am Chem Soc 2014;136:13959–62. https://doi.org/10.1021/ja5032776.
[139] Bogatyreva NS, Osypov AA, Ivankov DN. KineticDB: a database of protein folding kinetics. Nucleic Acids Res 2009;37:342–6. https://doi.org/10.1093/nar/gkn696.
[140] Lane TJ, Pande VS. Inferring the rate-length law of protein folding. PLoS One 2013; 8:8–12. https://doi.org/10.1371/journal.pone.0078606.
[141] Piana S, Lindorff-Larsen K, Shaw DE. Atomic-level description of ubiquitin folding. Proc Natl Acad Sci 2013;110:5915–20. https://doi.org/10.1073/pnas.1218321110.
[142] Liwo A, Arlukowicz P, Czaplewski C, Oldziej S, Pillardy J, Scheraga HA. A method for optimizing potential-energy functions by a hierarchical design of the potential-energy landscape: application to the UNRES force field. Proc Natl Acad Sci 2002; 99:1937–42. https://doi.org/10.1073/pnas.032675399.
[143] Tai C-H, Bai H, Taylor TJ, Lee B. Assessment of template-free modeling in CASP10 and ROLL. Proteins Struct Funct Bioinforma 2014;82:57–83. https://doi.org/10.1002/prot.24470.
[144] He Y, Mozolewska MA, Krupa P, Sieradzan AK, Wirecki TK, Liwo A, et al. Lessons from application of the UNRES force field to predictions of structures of CASP10 targets. Proc Natl Acad Sci 2013;110:14936–41. https://doi.org/10.1073/pnas.1313316110.
[145] Gront D, Kmiecik S, Blaszczyk M, Ekonomiuk D, Koliński A. Optimization of protein models. Wiley Interdiscip Rev Comput Mol Sci 2012;2:479–93. https://doi.org/10.1002/wcms.1090.
[146] Nugent T, Cozzetto D, Jones DT. Evaluation of predictions in the CASP10 model refinement category. Proteins 2014;82(Suppl. 2):98–111. https://doi.org/10.1002/prot.24377.
[147] Simmerling C, Lee MR, Ortiz AR, Kolinski A, Skolnick J, Kollman PA. Combining MONSSTER and LES/PME to Predict Protein Structure from Amino acid sequence: application to the small Protein CMTI-1. J Am Chem Soc 2000;122:8392–402. https://doi.org/10.1021/ja993119k.
[148] Lee MR, Baker D, Kollman PA. 2.1 and 1.8 Å average C α RMSD structure predictions on two small Proteins, HP-36 and S15. J Am Chem Soc 2001;123:1040–6. https://doi.org/10.1021/ja003150i.
[149] Lee MR, Tsai J, Baker D, Kollman PA. Molecular dynamics in the endgame of protein structure prediction. J Mol Biol 2001;313:417–30. https://doi.org/10.1006/jmbi.2001.5032.
[150] Chopra G, Summa CM, Levitt M. Solvent dramatically affects protein structure refinement. Proc Natl Acad Sci U S A 2008;105:20239–44. https://doi.org/10.1073/pnas.0810818105.
[151] Stumpff-Kane AW, Maksimiak K, Lee MS, Feig M. Sampling of near-native protein conformations during protein structure refinement using a coarse-grained model, normal modes, and molecular dynamics simulations. Proteins Struct Funct Bioinforma 2007;70:1345–56. https://doi.org/10.1002/prot.21674.
[152] Zhu J, Fan H, Periole X, Honig B, Mark AE. Refining homology models by combining replica-exchange molecular dynamics and statistical potentials. Proteins Struct Funct Bioinforma 2008;72:1171–88. https://doi.org/10.1002/prot.22005.
[153] Fan H, Periole X, Mark AE. Mimicking the action of folding chaperones by Hamiltonian replica-exchange molecular dynamics simulations: application in the refinement of de novo models. Proteins 2012;80:1744–54. https://doi.org/10.1002/prot.24068.
[154] Chen J, Brooks CL. Can molecular dynamics simulations provide high-resolution refinement of protein structure? Proteins Struct Funct Bioinforma 2007;67:922–30. https://doi.org/10.1002/prot.21345.
[155] Mirjalili V, Feig M. Protein structure refinement through structure selection and averaging from molecular dynamics ensembles. J Chem Theory Comput 2013;9: 1294–303. https://doi.org/10.1021/ct300962x.
[156] Mirjalili V, Noyes K, Feig M. Physics-based protein structure refinement through multiple molecular dynamics trajectories and structure averaging. Proteins Struct Funct Bioinforma 2014;82:196–207. https://doi.org/10.1002/prot.24336.
[157] Zhou H, Zhou Y. Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction. Protein Sci 2009;11:2714–26. https://doi.org/10.1110/ps.0217002.
[158] Cheng Q, Joung I, Lee J. A simple and efficient Protein structure refinement method. J Chem Theory Comput 2017;13:5146–62. https://doi.org/10.1021/acs.jctc.7b00470.
[159] Hovan L, Oleinikovas V, Yalinca H, Kryshtafovych A, Saladino G, Gervasio FL. Assessment of the model refinement category in CASP12. Proteins Struct Funct Bioinforma 2018;86:152–67. https://doi.org/10.1002/prot.25409.
[160] Raval A, Piana S, Eastwood MP, Dror RO, Shaw DE. Refinement of protein structure homology models via long, all-atom molecular dynamics simulations. Proteins 2012;80:2071–9. https://doi.org/10.1002/prot.24098.
[161] Xun S, Jiang F, Wu Y-D. Significant refinement of Protein structure models Using a Residue-Specific force field. J Chem Theory Comput 2015;11:1949–56. https://doi.org/10.1021/acs.jctc.5b00029.
[162] Heo L, Feig M. What makes it difficult to refine protein models further via molecular dynamics simulations? Proteins Struct Funct Bioinforma 2018;86:177–88. https://doi.org/10.1002/prot.25393.
[163] Raval A, Piana S, Eastwood MP, Shaw DE. Assessment of the utility of contact-based restraints in accelerating the prediction of protein structure using molecular dynamics simulations. Protein Sci 2016;25:19–29. https://doi.org/10.1002/pro.2770.
[164] Ogorzalek TL, Hura GL, Belsom A, Burnett KH, Kryshtafovych A, Tainer JA, et al. Small angle X-ray scattering and cross-linking for data assisted protein structure prediction in CASP 12 with prospects for improved accuracy. Proteins Struct Funct Bioinforma 2018;86:202–14. https://doi.org/10.1002/prot.25452.
[165] Wolynes PG, Friedrichs MS. Toward Protein tertiary structure recognition by means of associative memory Hamiltonians. Science 1989;246:371–3. https://doi.org/10.1126/science.246.4928.371.
[166] Friedrichs MS, Goldstein RA, Wolynes PG. Generalized protein tertiary structure recognition using associative memory hamiltonians. J Mol Biol 1991;222: 1013–34. https://doi.org/10.1016/0022-2836(91)90591-S.
[167] Hardin C, Eastwood MP, Luthey-Schulten Z, Wolynes PG. Associative memory Hamiltonians for structure prediction without homology: alpha-helical proteins. Proc Natl Acad Sci 2000;97:14235–40. https://doi.org/10.1073/pnas.230432197.
[168] Hardin C, Eastwood MP, Prentiss MC, Luthey-Schulten Z, Wolynes PG. Associative memory Hamiltonians for structure prediction without homology: alpha/beta proteins. Proc Natl Acad Sci U S A 2003;100:1679–84. https://doi.org/10.1073/pnas.252753899.
[169] Ulander J, Luthey-Schulten Z, Eastwood MP, Wolynes PG, Papoian GA. From the cover: water in protein structure prediction. Proc Natl Acad Sci 2004;101: 3352–7. https://doi.org/10.1073/pnas.0307851100.
[170] Davtyan A, Zheng W, Schafer NP, Wolynes PG, Papoian GA, Zheng W, et al. AWSEM-MD: protein structure prediction using coarse-grained physical potentials and bioinformatically based local structure biasing. J Phys Chem B 2012;116: 8494–503. https://doi.org/10.1016/j.bpj.2011.11.3373.
[171] Chen M, Lin X, Zheng W, Onuchic JN, Wolynes PG. Protein folding and structure prediction from the ground up: the atomistic associative memory, water mediated, structure and energy model. J Phys Chem B 2016;120:8557–65. https://doi.org/10.1021/acs.jpcb.6b02451.
[172] Karczyńska AS, Mozolewska MA, Krupa P, Giełdoń A, Liwo A, Czaplewski C. Prediction of protein structure with the coarse-grained UNRES force field assisted by small X-ray scattering data and knowledge-based information. Proteins Struct Funct Bioinforma 2018;86:228–39. https://doi.org/10.1002/prot.25421.
[173] Pilla KB, Gaalswyk K, MacCallum JL. Molecular modeling of biomolecules by paramagnetic NMR and computational hybrid methods. Biochim Biophys Acta Proteins Proteomics 1865;2017:1654–63. https://doi.org/10.1016/j.bbapap.2017.06.016.
[174] Brodie NI, Popov KI, Petrotchenko EV, Dokholyan NV, Borchers CH. Solving protein structures using short-distance cross-linking constraints as a guide for discrete molecular dynamics simulations. Sci Adv 2017;3:e1700479. https://doi.org/10.1126/sciadv.1700479.
[175] Merkley ED, Rysavy S, Kahraman A, Hafen RP, Daggett V, Adkins JN. Distance restraints from crosslinking mass spectrometry: mining a molecular dynamics simulation database to evaluate lysine-lysine distances. Protein Sci 2014;23:747–59. https://doi.org/10.1002/pro.2458.
[176] MacCallum JL, Perez A, Dill KA. Determining protein structures by combining semireliable data with atomistic physical models by Bayesian inference. Proc Natl Acad Sci U S A 2015;112:6985–90. https://doi.org/10.1073/pnas.1506788112.
[177] Perez A, Morrone JA, Dill KA. Accelerating physical simulations of proteins by leveraging external knowledge. Wiley Interdiscip Rev Comput Mol Sci 2017;7: e1309. https://doi.org/10.1002/wcms.1309.
[178] Perez A, MacCallum JL, Dill KA. Accelerating molecular simulations of proteins using Bayesian inference on weak information. Proc Natl Acad Sci 2015;112: 11846–51. https://doi.org/10.1073/pnas.1515561112.
[179] Perez A, Morrone JA, Brini E, MacCallum JL, Dill KA. Blind protein structure prediction using accelerated free-energy simulations. Sci Adv 2016;2:e1601274. https://doi.org/10.1126/sciadv.1601274.
[180] Peng J, Zhang Z. Unraveling low-resolution structural data of large biomolecules by constructing atomic models with experiment-targeted parallel cascade selection simulations. Sci Rep 2016;6:29360. https://doi.org/10.1038/srep29360.
[181] Mittal S, Shukla D. Recruiting machine learning methods for molecular simulations of proteins. Mol Simul 2018;44:891–904. https://doi.org/10.1080/08927022.2018.1448976.