Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Assessing the accuracy of physical models used in protein-folding simulations: quantitative evidence from long molecular dynamics simulations☆

Stefano Piana[1], John L Klepeis[1] and David E Shaw[1,2]

Advances in computer hardware, software and algorithms have now made it possible to run atomistically detailed, physics-based molecular dynamics simulations of sufficient length to observe multiple instances of protein folding and unfolding within a single equilibrium trajectory. Although such studies have already begun to provide new insights into the process of protein folding, realizing the full potential of this approach will depend not only on simulation speed, but on the accuracy of the physical models ('force fields') on which such simulations are based. While experimental data are not available for comparison with all of the salient characteristics observable in long protein-folding simulations, we examine here the extent to which current force fields reproduce (and fail to reproduce) certain relevant properties for which such comparisons are possible.

Addresses
[1] D. E. Shaw Research, New York, NY 10036, USA
[2] Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY 10032, USA

Corresponding authors: Piana, Stefano (Stefano.Piana-Agostinetti@DEShawResearch.com), Shaw, David E (David.Shaw@DEShawResearch.com)

Current Opinion in Structural Biology 2014, **24**:98–105

This review comes from a themed issue on **Folding and binding**

Edited by **James CA Bardwell** and **Gideon Schreiber**

For a complete overview see the Issue and the Editorial

Available online 24th January 2014

0959-440X/$ – see front matter, © 2013 The Authors. Published by Elsevier Ltd.

http://dx.doi.org/10.1016/j.sbi.2013.12.006

Introduction

A longstanding goal within the field of molecular biology has been to understand the principles that govern protein folding — the self-assembly process that leads from an unstructured polypeptide chain to a fully functional protein [1–5]. Although important progress toward this goal has been made using various experimental techniques [6•,7,8], such methods do not permit the direct examination of complete, continuous folding pathways at an atomistic level of detail. Molecular dynamics (MD) simulations, on the other hand, generate continuous, atomic-resolution trajectories, providing a potentially powerful complement to experimental results in elucidating key aspects of the folding process. Such simulations, however, are based on inexact models (force fields) of the forces underlying protein dynamics, and are also extremely demanding from a computational viewpoint, thus limiting the duration of the biological phenomena that may in practice be simulated.

Although simulations have been used for decades to study various aspects of the protein-folding process [9], until recently the length of a typical all-atom, physics-based MD simulation fell short of the folding times of even the fastest-folding proteins. Researchers often employed alternative approaches and approximations that were less computationally demanding than conventional MD, but capable of shedding light on at least some salient characteristics of the folding process [10•,11,12,13•,14–29]. (Examples include simplified abstractions of the protein [30,31], implicit solvent models [32], enhanced sampling algorithms [33–35], and techniques involving the aggregation and analysis of large numbers of short simulations [36,37].) Relatively few attempts [38–40] were made, however, to observe complete folding trajectories using atomic-resolution MD simulations with physically realistic force fields.

The development of special-purpose hardware for high-speed MD simulations [41], however, has extended the reach of all-atom, physics-based, explicit-solvent simulations to periods on the order of a millisecond — about two orders of magnitude beyond what was previously feasible, and significantly longer than the folding times of many fast-folding proteins. Long simulations can be especially useful when performed near the protein melting temperature, where the folded and unfolded states are equally populated and folding and unfolding occur on the same timescale. A single such MD simulation can encompass multiple folding and unfolding events [42], allowing a detailed mechanistic analysis of the folding process and meaningful estimates of various quantities related to the kinetics and thermodynamics of folding [43•].

These simulation studies have examined the folding of proteins up to ~100 amino acids in size [44], since larger proteins tend to fold on longer timescales. Using a simple

relation between chain length and folding time [45], we estimate that ~10% of the single-chain proteins in the Protein Data Bank (PDB) have folding times that fall within a range now accessible to MD simulation (a few milliseconds or less). This fraction is expected to grow as further advances are made in special-purpose hardware and algorithms for high-speed MD simulation, and may reach a third of the PDB within the next decade.

Since the folding times of many proteins now fall within a range accessible to MD simulation, the extent to which such simulations are able to extend our understanding of protein folding is becoming increasingly dependent on force field accuracy. On the one hand, it is now clear that current state-of-the-art force fields, particularly when employed in very long MD simulations, are sufficiently accurate to generate significant new insights into the folding process [46,47]. The physical models on which such force fields are based [48], however, are far from perfect. Understanding the nature and magnitude of the remaining discrepancies can be useful both in assessing the credibility of various types of simulation results and in modifying existing force fields to further improve their accuracy. In this review, we examine certain data derived from long MD simulations published over the last few years with an eye toward evaluating the accuracy of the force fields typically used in protein-folding studies.

While some tests of force field accuracy (e.g. comparisons with the results of quantum mechanical calculations) can be performed using only computational data, the extent to which biologically significant experimental findings are recapitulated in simulation provides a singularly important touchstone for evaluating the utility of a force field. This form of evaluation, however, is complicated by the very characteristic that makes simulations most valuable as a complement to experiments: their ability to provide information that is difficult to obtain experimentally. In particular, some of the most interesting findings derived from long MD simulations of protein folding involve the sequence of events that occur as proteins transition over their folding barriers. Direct experimental characterization of the folding-barrier region is difficult, however, due to its low equilibrium population, which limits the experimental data available for quantitative comparison with simulation results.

In light of this limitation, we focus in this review on a number of folding-related properties for which data are available both from experiments and from long-timescale, explicit-solvent MD simulations. We begin by examining the extent to which current force fields reproduce various experimentally measured properties of the two endpoints of the protein-folding process: the folded and unfolded states. We find that in many cases, simulation studies provide a remarkably accurate description of the structure and dynamics of the highly ordered folded states typical of globular proteins. The unfolded states observed in simulation, however, are typically more collapsed than those found experimentally. We then compare experimental and simulation-based measurements of certain global protein-folding observables like folding enthalpies and folding rates. We find that folding rates and folding free energies are in reasonable agreement with experiment, while protein-folding enthalpies are typically too small.

Modeling the folded state of proteins

Recent papers have critically evaluated the performance of force-fields commonly used in protein-folding simulations, focusing primarily on how accurately they reproduce the native-state structure and dynamics of proteins [49•,50•,51•,52,53]. Several of these force fields were found to provide rather accurate representations of the structure and dynamics of a number of small globular proteins on the submicrosecond timescale (although long simulations of larger, more flexible proteins were found to be more problematic [54]). In the latest variants of the Amber99SB-ILDN force field [55] (Amber99SB*-ILDN [56] and Amber99SB-ILDN-NMR [57]) and of the CHARMM force field (CHARMM22* [58] and CHARMM36 [59]), the error in the calculated NMR observables was often found to be similar to the experimental error.

Beauchamp and colleagues also examined the performance of these force fields in a variety of solvents and concluded that while the use of implicit solvation leads to severely degraded accuracy, there is not a large difference among the levels of accuracy achieved with different water models [51•]. A slightly different conclusion was reached by Cerutti and colleagues, who investigated the stability of a protein crystal using various combinations of protein force fields and water models [60]. They observed clear differences between force fields and found that three-point water models, which are recommended for use with Amber and CHARMM protein force fields, performed better than a four-point water model. In general, it was also found that Amber99SB gave the best results in terms of B-factors and stability of the protein lattice.

Because of the comprehensive nature of these types of studies, in which a number of protein systems must be simulated using a variety of different force fields, individual simulations have often been relatively short — typically less than 1 μs. Transitions among distinct conformational states of a folded protein occur relatively infrequently, however, and often go unobserved in submicrosecond simulations. As a result, simulations on that timescale are often able to sample only a region of the protein's conformational space that lies relatively close to its starting structure. This raises the question of whether the force field accuracy results obtained on a ns-to-μs

timescale are also observed in simulations of μs-to-ms duration — a timescale that often encompasses a significant number of state transitions, and a larger portion of the conformational space.

A study that employed simulations up to 10 μs long [50•] reported results in line with those obtained from shorter simulations: comparisons with experimental data revealed that several state-of-the-art force fields now appear to provide an accurate description of many structural and dynamical properties of proteins. Millisecond-scale simulations of the folded states of two small, globular proteins (BPTI [61] and ubiquitin [62•]) have provided an even broader, more demanding touchstone for the evaluation of force field accuracy. In both of these studies, transitions were observed among several alternative native-like structures that were not observed in shorter simulations. The experimentally determined native structure of both proteins appeared as one of the most populated conformations, and experimental support was also found for the existence of some of the alternative conformations.

Although there is room for further improvement, the extent to which these tests succeed in reproducing many native-state observables, often with experimental accuracy, suggests that at least in favorable cases, some of the force fields currently used in MD simulations of protein folding are sufficiently accurate to provide a remarkably faithful description of the folded state of small globular proteins.

Modeling the unfolded state of proteins

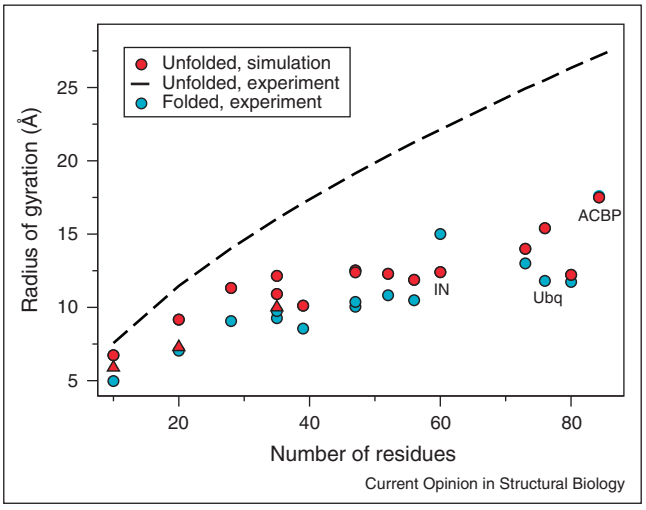
The unfolded state of a protein is characterized by a large number of different conformations with similar free energies, and even small force field inaccuracies can significantly alter the structural and dynamical properties of this state. Simulations of the unfolded state thus provide a potentially rich source of information about both the magnitude and origin of force field inaccuracies.

Among the most useful force-field properties that can be examined in unfolded-state simulations is the relative stability of different secondary structure elements. Non-physical imbalances in the tendency to form alternative secondary structures may impose significant limitations on the usefulness of computational studies of protein folding [63], and such imbalances are often more readily apparent in simulations of the unfolded state than in simulations of folded proteins. By way of example, the CHARMM27 force field, which provides an excellent description of the folded state of proteins, generates unfolded states that are substantially more helical than are found experimentally [50•]. Such imbalances can have significant consequences not only in studies of the unfolded state itself, but also in investigations that aim to elucidate various aspects of the protein-folding process [50•,63,64].

In the helix/coil-balanced variants of the Amber [56,65] and CHARMM [58,59] force fields, the problem of obtaining the correct stabilities for helical structures in the unfolded state is addressed by explicitly including in the parameterization NMR and CD data reporting the helical content of small alanine-based peptides [56]. These helix/coil-balanced force fields have been shown to provide a remarkably accurate description of the conformational preferences of short polypeptides [51•,56,59], and are among the most transferable in protein-folding simulations [43•,66,67]. Remaining discrepancies observed in small-peptide simulations are a slightly incorrect balance between the stability of PPII and β conformations [68] and an enthalpy of helix formation that is smaller than the experimental value [50•,56].

It has been observed that MD simulations of proteins larger than 20–30 amino acids tend to produce unfolded states that are more compact and structured than those suggested experimentally [69•]. By way of example, Figure 1 reports the radius of gyration (R_g) of the pH-denatured Acyl-CoA Binding Protein (ACBP); of the N-terminal domain of HIV-1 integrase (IN), which is natively unfolded in the absence of zinc; and of the

Figure 1



Comparison of experimental and simulation-derived radii of gyration. Red points represent the radius of gyration (R_g) calculated from MD simulations of a number of unfolded proteins, including the N-terminal domain of HIV-1 integrase (IN) [73], ubiquitin (Ubq) [74], and the pH-denatured Acyl-CoA binding protein (ACBP) [75]. (Data points marked with red circles are from simulations performed with the CHARMM22* force field [62•,66,72], while those marked with red squares are performed with the ff99SB*-ILDN force field [43•].) The dashed line represents the corresponding experimental R_g values for unfolded protein chains of the same length, as generated by the model of Kohn *et al.* [70], which was fitted based on experimental data from a large number of proteins under conditions of high denaturant concentration. The unfolded state observed in simulation is far more collapsed than the corresponding experimental values, and is in fact only slightly more expanded than the folded state (blue circles).

folded and unfolded states of 13 proteins simulated close to their respective melting temperatures. The R_g of the unfolded state of most of these proteins is only marginally larger than that of the folded state, and is much smaller than the R_g of proteins of similar sizes unfolded in solutions with a high denaturant concentration [70].

While in the absence of denaturant some degree of hydrophobic collapse is expected [71], this collapse is usually smaller than that observed in simulations. A direct comparison between experiment and simulation can be made for three proteins, whose R_g has been measured experimentally under conditions similar to the simulation conditions (IN [SP and DES, unpublished results]; ubiquitin [50•]; and ACBP [72]); in all three cases the experimental R_g (IN 23 Å [73]; ubiquitin 27 Å [74]; ACBP 19.5 Å [75]) is significantly larger than that observed in simulation (Figure 1). Similarly collapsed unfolded states were also obtained by other researchers studying other combinations of force fields and water models in simulations of the GB1 hairpin [17], CspTm [76], and Protein L [77].

The reason for these discrepancies is not entirely clear, but both the Amber and CHARMM force fields exhibit the same symptoms, with Amber (red squares in Figure 1) generally generating more compact unfolded states than CHARMM (red circles in Figure 1). These results suggest that there may be a general problem with either the functional form or the force-field parameterization. An abnormal collapse could alter the structural properties of the unfolded state, encouraging the formation of native and non-native secondary structure motifs (in particular, α helices) [74,78•], and considerably slowing down the dynamics [72,77].

Protein-folding rates and thermodynamics

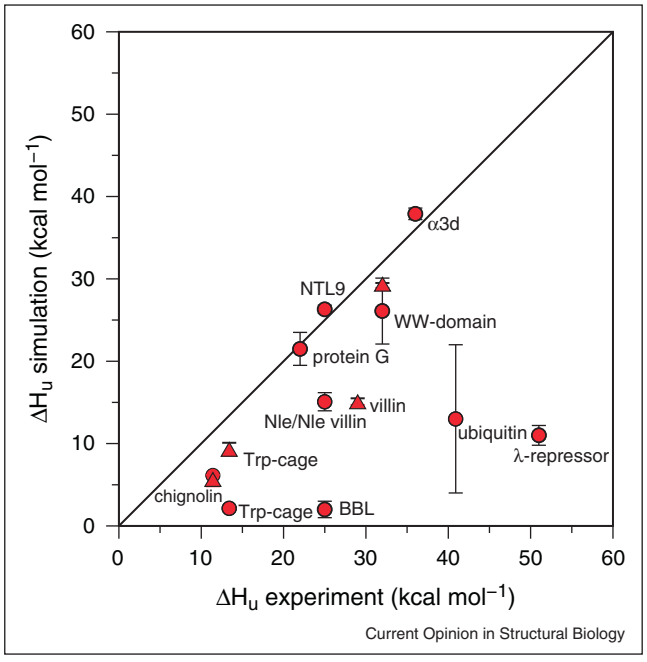
Despite the fact that some of the structural properties of the unfolded state are not well reproduced by current force fields (see section: 'Modeling the unfolded state of proteins'), good agreement has often been observed between the calculated and experimental values for folding rates and melting temperatures. In some cases, this may be the result of cancellation of errors. By way of example, protein-folding simulations performed using the TIP3P water model, which has a viscosity much lower than the experimentally determined value for real water [79], would be expected to result in calculated folding rates that are too fast, but the folding rates observed in simulation are actually equal to or slower than the corresponding experimental values [62•,66]. Although it is difficult to ascertain the cause of this result, explanations include the possibility that simulations may overestimate the folding free-energy barriers or the amount of internal friction present during structural rearrangements.

Calculated folding free energies are also often (though not always) highly accurate. While this suggests that in some cases protein-folding simulations can be successfully used to make quantitative predictions, care should be taken not to overinterpret these findings, since negative results obtained from unsuccessful folding simulations in which the native fold ultimately turns out to be unstable are rarely published (for a notable exception, see [40,64]). In a systematic attempt to study the folding of a number of fast-folding protein domains with simulations conducted using a single force field [66], the calculated melting temperature for most of the proteins investigated was found to be lower — sometimes by tens of degrees — than the experimentally estimated value, and for one of the proteins it was so low that folding simulations could not be successfully performed. It is worth noting that these results were obtained with CHARMM22*, the force field that exhibited the highest transferability among different protein classes.

Performing a direct experimental validation of the folding mechanisms observed in simulations requires an experimental characterization of the intermediate states of protein folding, which is a difficult task. Techniques such as Φ-value analysis and its variants [8,80], however, which are based on the comparison of folding rates and folding free energies of point mutants of a protein, can provide indirect structural information on the transition state for protein folding. Calculating Φ values from simulation is conceptually straightforward, but this approach can require a great deal of computer time, since accurate folding rates and folding free energies must be calculated not only for the wild-type protein, but also for multiple mutants, by performing extensive reversible folding simulations for each of them [42,81]. In the few cases in which such calculations have been performed using atomistic, explicit-solvent simulations [43•,61], good agreement has generally been observed between the calculated and experimental Φ values. These results indicate that force fields can capture the effect of various mutations on folding rates and folding stabilities, and lend some support to the folding mechanisms observed in such simulations.

While good agreement with experiment can generally be obtained for folding times and protein stabilities, large discrepancies are typically observed for thermodynamic properties like the enthalpy and heat capacity of folding. Folding enthalpies derived from simulations are generally much smaller than those obtained from experiments (Figure 2), especially for helical proteins (with the notable exception of the three-helix bundle α3d, whose calculated folding enthalpy is in excellent agreement with the experimental estimate). Heat capacities of folding are more difficult to calculate from simulation with high precision, but the available data suggest that they may also be too small [43•,66]. It has been suggested that part of this discrepancy may be due to the poor temperature-dependent properties of the TIP3P water model [82]

Figure 2



Comparison of experimental and simulation-derived unfolding enthalpies. In this scatter plot, circles represent values calculated using the CHARMM22* force field [62•,66], while triangles represent values calculated using Amber ff99SB*-ILDN [43•]. Experimental measurements for chignolin, Trp-cage, villin, and Nle/Nle villin were performed under conditions closely matching the simulation setup [92–96], while those of the other proteins were performed on slightly different protein sequences and/or at different pH [97–102].

used for many explicit-solvent simulations, and indeed, some improvement is observed in simulations performed with the TIP4P-2005 water model [65], which more accurately reproduces water properties across a wide temperature range [83].

Although the systematic underestimation of folding enthalpies in simulation could be the result of a specific interaction being poorly described by current force fields, it is not clear whether (or to what extent) this is in fact the case. One alternative explanation, for example, might be a more general phenomenon arising from differences in the degree of structural heterogeneity of the folded and unfolded states. We begin by noting that the folding enthalpy may be thought of as the average enthalpy of the tightly constrained conformational ensemble that constitutes the native state minus the average enthalpy of the structurally diverse conformations that together constitute the unfolded state. Let us now assume the existence of a 'perfect' force field that exactly reproduces the enthalpies of the folded and unfolded conformational ensembles (and thus the folding enthalpy), and consider how the simulation-derived enthalpies of these two states would be likely to change if small perturbations were made to parameters of that perfect force field.

Perturbing the perfect force field would be expected to shift the composition of the structurally heterogeneous unfolded ensemble in such a way as to increase the population of conformations whose enthalpies have decreased and decrease the population of conformations whose enthalpies have increased. Other things being equal, force field perturbation should thus tend to decrease the average enthalpy of the unfolded state. The folded ensemble, on other hand, has a high degree of structural homogeneity, and perturbation of the force field is likely to have similar effects on the enthalpies of the various conformations that are present in that ensemble. The population-shifting effect should thus be far less pronounced in the case of the folded state, resulting in a smaller reduction in the average enthalpy. Other things being equal, force field errors should thus be expected to result in a reduction of the folding enthalpy measured in simulation.

This observation suggests that it may be valuable to include folding enthalpy as one of the key experimental quantities one tries to reproduce in the course of force-field optimization. Some examples along these lines can be found in the literature (e.g. using as a target function the enthalpy of helix formation for small alanine-based polypeptides [58,65,84], or using force-field optimization protocols that attempt to maximize the gap between the total energy of the native state and that of any possible alternative conformation [85–88]).

Conclusions

Now that fully atomistic, physics-based MD simulations exceed the millisecond timescale, force-field accuracy largely determines the overall accuracy achievable in computational studies of many fast-folding proteins. We have thus examined here how accurately physics-based force fields can reproduce experimental quantities that are relevant to protein folding. We find that the prediction of native-state structures and folding rates appears to be more robust with respect to errors in the potential-energy function [89–91] than is the prediction of the detailed kinetics or that of the unfolded-state structural properties. It is generally observed that the enthalpy of the folded state in simulation is lower than that measured experimentally; we argue, however, that this need not be the result of a specific force-field deficiency, but could be the result of the accumulation of a number of small, unrelated errors in the various force-field parameters.

While improvements in the potential-energy function have previously been mostly benchmarked against reproducing the structure and dynamics of the folded state, we recommend that further improvements also be measured against the structural properties of disordered states and the temperature dependence of fold stability of a wide range of different protein folds. Such efforts would be

greatly aided by the availability of high-quality experimental data probing the structure of the unfolded-state ensemble under native conditions, as well as by continued advances in computer software and hardware that would further extend the timescales accessible to MD simulations, allowing computational folding studies of larger and larger portions of the protein space.

Acknowledgement

We thank Berkman Frank for editorial assistance.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Pande VS, Grosberg AY, Tanaka T, Rokhsar DS: **Pathways for protein folding: is a new view needed?** *Curr Opin Struct Biol* 1998, **8**:68-79.
2. Daggett V, Fersht AR: **Is there a unifying mechanism for protein folding?** *Trends Biochem Sci* 2003, **28**:18-25.
3. Onuchic JN, Wolynes PG: **Theory of protein folding**. *Curr Opin Struct Biol* 2004, **14**:70-75.
4. Krishna MM, Englander SW: **A unified mechanism for protein folding: predetermined pathways with optional errors**. *Protein Sci* 2007, **16**:449-464.
5. Karplus M: **Behind the folding funnel diagram**. *Nat Chem Biol* 2011, **17**:401-404.
6. • Chung HS, McHale K, Louis JM, Eaton WA: **Single-molecule fluorescence experiments determine protein folding transition path times**. *Science* 2012, **335**:981-984.

This paper represents an important step forward toward the direct experimental characterization of transition paths for protein folding under equilibrium conditions. The transition-path times for folding determined experimentally turn out to be in remarkable agreement with previous simulation studies.

7. Sadqi M, Fushman D, Muñoz V: **Atom-by-atom analysis of global downhill protein folding**. *Nature* 2006, **442**:317-321.
8. Matouschek A, Kellis JT Jr, Serrano L, Fersht AR: **Mapping the transition state and pathway of protein folding by protein engineering**. *Nature* 1989, **340**:122-126.
9. Levitt M, Warshel A: **Computer simulation of protein folding**. *Nature* 1975, **253**:694-698.
10. • Voelz VA, Jäger M, Yao S, Chen Y, Zhu L, Waldauer SA, Bowman GR, Friedrichs M, Bakajin O, Lapidus LJ, Weiss S, Pande VS: **Slow unfolded-state structuring in Acyl-CoA binding protein folding revealed by simulation and experiment**. *J Am Chem Soc* 2012, **134**:12565-12577.

High-temperature, implicit-solvent simulations and ultrafast mixing experiments suggest that the interconversion between different structures in the unfolded state may be slower than commonly believed.

11. Granata D, Camilloni C, Vendruscolo M, Laio A: **Characterization of the free-energy landscapes of proteins by NMR-guided metadynamics**. *Proc Natl Acad Sci U S A* 2013, **110**:6817-6822.
12. a Beccara S, Škrbić T, Covino R, Micheletti C, Faccioli P: **Folding pathways of a knotted protein with a realistic atomistic force field**. *PLoS Comput Biol* 2013, **9**:e1003002.
13. • Adhikari AN, Freed KF, Sosnick TR: **De novo prediction of protein folding pathways and structure using the principle of sequential stabilization**. *Proc Natl Acad Sci U S A* 2012, **109**:17442-17447.

(See also Ref [69•].) Starting from a model for the unfolded state based on coil libraries and a coarse-grained statistical potential, several rounds of simulation are performed, retaining the structures that are most often observed in each round. This leads to remarkably accurate predictions for the folded state and folding mechanism of proteins.

14. Shao Q, Gao YQ: **The relative helix and hydrogen bond stability in the B domain of protein A as revealed by integrated tempering sampling molecular dynamics simulation**. *J Chem Phys* 2011, **135**:135102.
15. Berteotti A, Barducci A, Parrinello M: **Effect of urea on the β-hairpin conformational ensemble and protein denaturation mechanism**. *J Am Chem Soc* 2011, **133**:17200-17206.
16. Paschek D, Day R, García AE: **Influence of water–protein hydrogen bonding on the stability of Trp-cage miniprotein. A comparison between the TIP3P and TIP4P-Ew water models**. *Phys Chem Chem Phys* 2011, **13**:19840-19847.
17. Best RB, Mittal J: **Free-energy landscape of the GB1 hairpin in all-atom explicit solvent simulations with different force fields: similarities and differences**. *Proteins* 2011, **79**:1318-1328.
18. Mohanty S, Meinke JH, Zimmermann O: **Folding of Top7 in unbiased all-atom Monte Carlo simulations**. *Proteins* 2013, **81**:1446-1456.
19. Berhanu WM, Jiang P, Hansmann UH: **Folding and association of a homotetrameric protein complex in an all-atom Go model**. *Phys Rev E* 2013, **87**:014701.
20. Oshaben KM, Salari R, McCaslin DR, Chong LT, Horne WS: **The native GCN4 leucine-zipper domain does not uniquely specify a dimeric oligomerization state**. *Biochemistry* 2012, **51**:9581-9591.
21. Shao Q, Shi J, Zhu W: **Enhanced sampling molecular dynamics simulation captures experimentally suggested intermediate and unfolded states in the folding pathway of Trp-cage miniprotein**. *J Chem Phys* 2012, **137**:125103.
22. Marino KA, Bolhuis PG: **Confinement-induced states in the folding landscape of the Trp-cage miniprotein**. *J Phys Chem B* 2012, **116**:11872-11880.
23. Sułkowska JI, Noel JK, Onuchic JN: **Energy landscape of knotted protein folding**. *Proc Natl Acad Sci U S A* 2012, **109**:17783-17788.
24. Liu Y, Kellogg E, Liang H: **Canonical and micro-canonical analysis of folding of trpzip2: an all-atom replica exchange Monte Carlo simulation study**. *J Chem Phys* 2012, **137**:045103.
25. Kührová P, De Simone A, Otyepka M, Best RB: **Force-field dependence of chignolin folding and misfolding: comparison with experiment and redesign**. *Biophys J* 2012, **102**:1897-1906.
26. Okumura H: **Temperature and pressure denaturation of chignolin: folding and unfolding simulation by multibaric–multithermal molecular dynamics method**. *Proteins* 2012, **80**:2397-2416.
27. Zhang C, Ma J: **Folding helical proteins in explicit solvent using dihedral-biased tempering**. *Proc Natl Acad Sci U S A* 2012, **109**:8139-8144.
28. Maisuradze GG, Zhou R, Liwo A, Xiao Y, Scheraga HA: **Effects of mutation, truncation, and temperature on the folding kinetics of a WW domain**. *J Mol Biol* 2012, **420**:350-365.
29. Kouza M, Hansmann UH: **Folding simulations of the A and B domains of protein G**. *J Phys Chem B* 2012, **116**:6645-6653.
30. Hyeon C, Thirumalai D: **Capturing the essence of folding and functions of biomolecules using coarse-grained models**. *Nat Commun* 2011, **2**:487.
31. Riniker S, Allison JR, van Gunsteren WF: **On developing coarse-grained models for biomolecular simulation: a review**. *Phys Chem Chem Phys* 2012, **14**:12423-12430.
32. Chen J, Brooks CL 3rd: **Implicit modeling of nonpolar solvation for simulating protein folding and conformational transitions**. *Phys Chem Chem Phys* 2008, **10**:471-481.
33. Mitsutake A, Mori Y, Okamoto Y: **Enhanced sampling algorithms**. *Methods Mol Biol* 2013, **924**:153-195.
34. Leone V, Marinelli F, Carloni P, Parrinello M: **Targeting biomolecular flexibility with metadynamics**. *Curr Opin Struct Biol* 2010, **20**:148-154.

35. Christen M, van Gunsteren WF: **On searching in, sampling of, and dynamically moving through conformational space of biomolecular systems: a review**. *J Comput Chem* 2008, **29**: 157-166.

36. Noé F, Fischer S: **Transition networks for modeling the kinetics of conformational change in macromolecules**. *Curr Opin Struct Biol* 2008, **18**:154-162.

37. Lane TJ, Shukla D, Beauchamp KA, Pande VS: **To milliseconds and beyond: challenges in the simulation of protein folding**. *Curr Opin Struct Biol* 2013, **23**:58-65.

38. Duan Y, Kollman PA: **Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution**. *Science* 1998, **282**:740-744.

39. 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**:173-183.

40. Freddolino PL, Liu F, Gruebele M, Schulten K: **Ten-microsecond molecular dynamics simulation of a fast-folding WW domain**. *Biophys J* 2008, **94**:L75-L77.

41. Shaw DE, Dror RO, Salmon JK, Grossman JP, Mackenzie KM, Bank JA, Young C, Deneroff MM, Batson B, Bowers KJ, Chow E, Eastwood MP, Ierardi DJ, Klepeis JL, Kuskin JS, Larson RH, Lindorff-Larsen K, Maragakis P, Moraes MA, Piana S, Shan Y, Towles B: **Millisecond-scale molecular dynamics simulations on Anton**. In *Proceedings of the Conference on High Performance Computing, Networking, Storage and Analysis (SC09)*. New York, NY: ACM; 2009, .

42. Ferrara P, Caflisch A: **Folding simulations of a three-stranded antiparallel beta-sheet peptide**. *Proc Natl Acad Sci U S A* 2000, **97**:10780-10785.

43. • Piana S, Lindorff-Larsen K, Shaw DE: **Protein folding kinetics and thermodynamics from atomistic simulation**. *Proc Natl Acad Sci U S A* 2012, **109**:17845-17850.
Reversible simulations of several variants of the villin headpiece at multiple temperatures are used to estimate thermodynamic and kinetic quantities using various approaches. A Φ-value analysis indicates that the Nle/Nle double mutant folds through a slightly different pathway than the wild type.

44. Piana S, Lindorff-Larsen K, Shaw DE: **Atomistic description of the folding of a dimeric protein**. *J Phys Chem B* 2013, **117**:12935-12942.

45. Ivankov DN, Garbuzynskiy SO, Alm E, Plaxco KW, Baker D, Finkelstein AV: **Contact order revisited: influence of protein size on the folding rate**. *Protein Sci* 2003, **12**:2057-2062.

46. Best RB: **Atomistic molecular simulations of protein folding**. *Curr Opin Struct Biol* 2012, **22**:52-61.

47. Prigozhin MB, Gruebele M: **Microsecond folding experiments and simulations: a match is made**. *Phys Chem Chem Phys* 2013, **15**:3372-3388.

48. Guvench O, MacKerell AD Jr: **Comparison of protein force fields for molecular dynamics simulations**. *Methods Mol Biol* 2008, **443**:63-88.

49. • Huang J, MacKerell AD Jr: **CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data**. *J Comput Chem* 2013, **34**:2135-2145.
The CHARMM36 force field includes corrections to the side-chain and backbone-torsion potentials based on both QM data and small-peptide simulations. It is shown that this force field represents a substantial improvement with respect to previous CHARMM versions.

50. • Lindorff-Larsen K, Maragakis P, Piana S, Eastwood MP, Dror RO, Shaw DE: **Systematic validation of protein force fields against experimental data**. *PLoS ONE* 2012, **7**:e32131.
A range of force fields are evaluated with regard to protein folding, their ability to reproduce the folded state of proteins, and the conformational ensemble of small peptides. It is shown that protein force fields are markedly improving with time.

51. • Beauchamp KA, Lin YS, Das R, Pande VS: **Are protein force fields getting better? A systematic benchmark on 524 diverse NMR measurements**. *J Chem Theory Comput* 2012, **8**:1409-1414.
A number of combinations of water models and protein force fields are systematically evaluated in their ability to reproduce NMR measurements.

52. Patapati KK, Glykos NM: **Three force fields' views of the 3(10) helix**. *Biophys J* 2011, **101**:1766-1771.

53. Lange OF, van der Spoel D, de Groot BL: **Scrutinizing molecular mechanics force fields on the submicrosecond timescale with NMR data**. *Biophys J* 2010, **99**:647-655.

54. 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**:2071-2079.

55. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, Shaw DE: **Improved side-chain torsion potentials for the Amber ff99SB protein force field**. *Proteins* 2010, **78**:1950-1958.

56. Best RB, Hummer G: **Optimized molecular dynamics force fields applied to the helix–coil transition of polypeptides**. *J Phys Chem B* 2009, **113**:9004-9015.

57. Li DW, Brüschweiler R: **NMR-based protein potentials**. *Angew Chem* 2010, **122**:6930-6932.

58. Piana S, Lindorff-Larsen K, Shaw DE: **How robust are protein folding simulations with respect to force field parameterization?** *Biophys J* 2011, **100**:L47-L49.

59. Best RB, Zhu X, Shim J, Lopes PE, Mittal J, Feig M, Mackerell AD Jr: **Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ and side-chain $\chi(1)$ and $\chi(2)$ dihedral angles**. *J Chem Theory Comput* 2012, **8**:3257-3273.

60. Cerutti DS, Freddolino PL, Duke RE Jr, Case DA: **Simulations of a protein crystal with a high resolution X-ray structure: evaluation of force fields and water models**. *J Phys Chem B* 2010, **114**:12811-12824.

61. Shaw DE, Maragakis P, Lindorff-Larsen K, Piana S, Dror RO, Eastwood MP, Bank JA, Jumper JM, Salmon JK, Shan Y, Wriggers W: **Atomic-level characterization of the structural dynamics of proteins**. *Science* 2010, **330**:341-346.

62. • Piana S, Lindorff-Larsen K, Shaw DE: **Atomic-level description of ubiquitin folding**. *Proc Natl Acad Sci U S A* 2013, **110**:5915-5920.
Millisecond-timescale equilibrium MD simulations are used to describe the folding kinetics and energy landscape of ubiquitin. The general features of the folding mechanism are found to be similar to those of fast-folding proteins.

63. Freddolino PL, Harrison CB, Liu Y, Schulten K: **Challenges in protein folding simulations: timescale, representation, and analysis**. *Nat Phys* 2010, **6**:751-758.

64. Freddolino PL, Park S, Roux B, Schulten K: **Force field bias in protein folding simulations**. *Biophys J* 2009, **96**:3772-3780.

65. Best RB, Mittal J: **Protein simulations with an optimized water model: cooperative helix formation and temperature-induced unfolded state collapse**. *J Phys Chem B* 2010, **114**:14916-14923.

66. Lindorff-Larsen K, Piana S, Dror RO, Shaw DE: **How fast-folding proteins fold**. *Science* 2011, **334**:517-520.

67. Mittal J, Best RB: **Tackling force-field bias in protein folding simulations: folding of villin HP35 and Pin WW domains in explicit water**. *Biophys J* 2010, **99**:L26-L28.

68. Verbaro D, Ghosh I, Nau WM, Schweitzer-Stenner R: **Discrepancies between conformational distributions of a polyalanine peptide in solution obtained from molecular dynamics force fields and amide I′ band profiles**. *J Phys Chem B* 2010, **114**:17201-17208.

69. • Adhikari AN, Freed KF, Sosnick TR: **Simplified protein models: predicting folding pathways and structure using amino acid sequences**. *Phys Rev Lett* 2013, **111**:028103.
See Ref [13•].

70. Kohn JE, Millett IS, Jacob J, Zagrovic B, Dillon TM, Cingel N, Dothager RS, Seifert S, Thiyagarajan P, Sosnick TR, Hasan MZ, Pande VS, Ruczinski I, Doniach S, Plaxco KW: **Random-coil behavior and the dimensions of chemically unfolded proteins**. *Proc Natl Acad Sci U S A* 2004, **101**:12491-12496.

71. Haran G: **How, when and why proteins collapse: the relation to folding**. *Curr Opin Struct Biol* 2012, **22**:14-20.

72. Lindorff-Larsen K, Trbovic N, Maragakis P, Piana S, Shaw DE: **Structure and dynamics of an unfolded protein examined by molecular dynamics simulation**. *J Am Chem Soc* 2012, **134**:3787-3791.

73. Müller-Späth S, Soranno A, Hirschfeld V, Hofmann H, Rüegger S, Reymond L, Nettels D, Schuler B: **Charge interactions can dominate the dimensions of intrinsically disordered proteins**. *Proc Natl Acad Sci U S A* 2010, **107**:14609-14614.

74. Jacob J, Krantz B, Dothager RS, Thiyagarajan P, Sosnick TR: **Early collapse is not an obligate step in protein folding**. *J Mol Biol* 2004, **338**:369-382.

75. Fieber W, Kragelund BB, Meldal M, Poulsen FM: **Reversible dimerization of acid-denatured ACBP controlled by helix A4**. *Biochemistry* 2005, **44**:1375-1384.

76. Nettels D, Müller-Späth S, Küster F, Hofmann H, Haenni D, Rüegger S, Reymond L, Hoffmann A, Kubelka J, Heinz B, Gast K, Best RB, Schuler B: **Single-molecule spectroscopy of the temperature-induced collapse of unfolded proteins**. *Proc Natl Acad Sci U S A* 2009, **106**:20740-20745.

77. Voelz VA, Singh VR, Wedemeyer WJ, Lapidus LJ, Pande VS: **Unfolded-state dynamics and structure of protein L characterized by simulation and experiment**. *J Am Chem Soc* 2010, **132**:4702-4709.

78. • Piana S, Lindorff-Larsen K, Dirks RM, Salmon JK, Dror RO, Shaw DE: **Evaluating the effects of cutoffs and treatment of long-range electrostatics in protein folding simulations**. *PLoS ONE* 2012, **7**:e39918.
In contrast with folding free energies and folding rates, unfolded-state structural properties can be highly sensitive to even modest approximations.

79. Mahoney MW, Jorgensen WL: **Diffusion constant of the TIP5P model of liquid water**. *J Chem Phys* 2001, **114**:363-366.

80. Krantz BA, Dothager RS, Sosnick TR: **Discerning the structure and energy of multiple transition states in protein folding using psi-analysis**. *J Mol Biol* 2004, **337**:463-475.

81. Settanni G, Rao F, Caflisch A: **Phi-value analysis by molecular dynamics simulations of reversible folding**. *Proc Natl Acad Sci U S A* 2005, **102**:628-633.

82. Jorgensen WL, Jenson C: **Temperature dependence of TIP3P, SPC, and TIP4P water from NPT Monte Carlo simulations: seeking temperatures of maximum density**. *J Comput Chem* 1998, **19**:1179-1186.

83. Abascal JL, Vega C: **A general purpose model for the condensed phases of water: TIP4P/2005**. *J Chem Phys* 2005, **123**:234505.

84. Best RB, Mittal J, Feig M, MacKerell AD Jr: **Inclusion of many-body effects in the additive CHARMM protein CMAP potential results in enhanced cooperativity of α-helix and β-hairpin formation**. *Biophys J* 2012, **103**:1045-1051.

85. Mirny LA, Shakhnovich EI: **How to derive a protein folding potential? A new approach to an old problem**. *J Mol Biol* 1996, **264**:1164-1179.

86. Goldstein RA, Luthey-Schulten ZA, Wolynes PG: **Optimal protein-folding codes from spin-glass theory**. *Proc Natl Acad Sci U S A* 1992, **89**:4918-4922.

87. Maiorov VN, Crippen GM: **Contact potential that recognizes the correct folding of globular proteins**. *J Mol Biol* 1992, **227**:876-888.

88. Vendruscolo M, Mirny LA, Shakhnovich EI, Domany E: **Comparison of two optimization methods to derive energy parameters for protein folding: perceptron and *Z* score**. *Proteins* 2000, **41**:192-201.

89. Bryngelson JD: **When is a potential accurate enough for structure prediction? Theory and application to a random heteropolymer model of protein folding**. *J Chem Phys* 1994, **100**:6038-6045.

90. Pande VS, Grosberg AY, Tanaka T: **How accurate must potentials be for successful modeling of protein folding?** *J Chem Phys* 1995, **103**:9482-9491.

91. Faver JC, Benson ML, He X, Roberts BP, Wang B, Marshall MS, Sherrill CD, Merz KM Jr: **The energy computation paradox and ab initio protein folding**. *PLoS ONE* 2011, **6**:e18868.

92. Chung HS, Shandiz A, Sosnick TR, Tokmakoff A: **Probing the folding transition state of ubiquitin mutants by temperature-jump-induced downhill unfolding**. *Biochemistry* 2008, **47**:13870-13877.

93. Honda S, Akiba T, Kato YS, Sawada Y, Sekijima M, Ishimura M, Ooishi A, Watanabe H, Odahara T, Harata K: **Crystal structure of a ten-amino acid protein**. *J Am Chem Soc* 2008, **130**: 15327-15331.

94. Barua B, Lin JC, Williams VD, Kummler P, Neidigh JW, Andersen NH: **The Trp-cage: optimizing the stability of a globular miniprotein**. *Protein Eng Des Sel* 2008, **21**:171-185.

95. Kubelka J, Chiu TK, Davies DR, Eaton WA, Hofrichter J: **Sub-microsecond protein folding**. *J Mol Biol* 2006, **359**:546-553.

96. Godoy-Ruiz R, Henry ER, Kubelka J, Hofrichter J, Muñoz V, Sanchez-Ruiz JM, Eaton WA: **Estimating free-energy barrier heights for an ultrafast folding protein from calorimetric and kinetic data**. *J Phys Chem B* 2008, **112**:5938-5949.

97. Cobos ES, Iglesias-Bexiga M, Ruiz-Sanz J, Mateo PL, Luque I, Martinez JC: **Thermodynamic characterization of the folding equilibrium of the human Nedd4-WW4 domain: at the frontiers of cooperative folding**. *Biochemistry* 2009, **48**:8712-8720.

98. Horng JC, Moroz V, Raleigh DP: **Rapid cooperative two-state folding of a miniature alpha–beta protein and design of a thermostable variant**. *J Mol Biol* 2003, **326**:1261-1270.

99. Neuweiler H, Sharpe TD, Rutherford TJ, Johnson CM, Allen MD, Ferguson N, Fersht AR: **The folding mechanism of BBL: plasticity of transition-state structure observed within an ultrafast folding protein family**. *J Mol Biol* 2009, **390**:1060-1073.

100. Alexander P, Orban J, Bryan P: **Kinetic analysis of folding and unfolding the 56 amino acid IgG-binding domain of streptococcal protein G**. *Biochemistry* 1992, **31**:7243-7248.

101. Zhu Y, Alonso DO, Maki K, Huang CY, Lahr SJ, Daggett V, Roder H, DeGrado WF, Gai F: **Ultrafast folding of alpha3D: a de novo designed three-helix bundle protein**. *Proc Natl Acad Sci U S A* 2003, **100**:15486-15491.

102. Merabet E, Ackers GK: **Calorimetric analysis of lambda cI repressor binding to DNA operator sites**. *Biochemistry* 1995, **34**:8554-8563.