ELSEVIER

The protein folding problem: when will it be solved?

Ken A Dill[1], S Banu Ozkan[2], Thomas R Weikl[3], John D Chodera[4] and Vincent A Voelz[4]

The protein folding problem can be viewed as three different problems: defining the thermodynamic folding code; devising a good computational structure prediction algorithm; and answering Levinthal's question regarding the kinetic mechanism of how proteins can fold so quickly. Once regarded as a grand challenge, protein folding has seen much progress in recent years. Folding codes are now being used to successfully design proteins and non-biological foldable polymers; aided by the Critical Assessment of Techniques for Structure Prediction (CASP) competition, protein structure prediction has now become quite good. Even the once-challenging Levinthal puzzle now seems to have an answer — a protein can avoid searching irrelevant conformations and fold quickly by making local independent decisions first, followed by non-local global decisions later.

Addresses
[1] Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94143, USA
[2] Department of Physics, Arizona State University, Tempe, AZ 85287, USA
[3] Max Planck Institute of Colloids and Interfaces, Department of Theory and Biosystems, 14424 Potsdam, Germany
[4] Graduate Group in Biophysics, University of California, San Francisco, CA 94143, USA

Corresponding author: Dill, Ken A (dill@maxwell.ucsf.edu)

Current Opinion in Structural Biology 2007, **17**:342–346

This review comes from a themed issue on
Sequences and topology
Edited by William R Pearson and Anna Tramontano

Available online 14th June 2007

0959-440X/$ – see front matter

DOI 10.1016/j.sbi.2007.06.001

Introduction

The amino acid sequence of a protein determines its structure, which determines its mechanism of action. This key paradigm in biochemistry accounts for nearly one in four Nobel Prizes in Chemistry since 1956 [1]. The protein folding problem is the question of how the amino acid sequence of a protein dictates its structure. We were asked to say when the folding problem would be solved. The general perception has been that the protein folding problem is a grand challenge that will require many supercomputer years to solve. For example, in 2005, *Science* named the protein folding problem one of the 125 biggest unsolved problems in science [2]. We argue here, instead, that great headway has been made, both theoretical and experimental, and that the central problems of principle, and even key problems of implementation and practice, have already been solved. We summarize progress, problems and new directions, as we see them.

Three problems of protein folding

The protein folding problem is three different problems: the folding code — the thermodynamic question of how a native structure results from the interatomic forces acting on an amino acid sequence; protein structure prediction — the computational problem of how to predict the native structure of a protein from its amino acid sequence; and folding speed (Levinthal's paradox) — the kinetic question of how a protein can fold so fast (the grand challenge noted above).

The folding code

Before the mid-1980s, the predominant view was that the protein folding code is the sum of many different small interactions (hydrogen bonds, ion pairs, van der Waals interactions, hydrophobic interactions), mainly expressed through secondary structures and mainly local in the sequence (i.e. near neighbors along the chain; see, for example, the review by Anfinsen and Scheraga [3]). However, through statistical mechanical modeling, a different view emerged in the late 1980s — namely, that there is a dominant component to the folding code (the hydrophobic interaction), that the folding code is distributed both locally and non-locally in the sequence, and that native secondary structures are more a consequence than a cause of folding forces [4].

Although there are alternative viewpoints [5], there are now many experiments showing that 'reduced-alphabet solvation-based' codes correctly encode native structures [6,7] and the amyloid-like aggregates that are formed by particular sequences [8••]. In addition, such codes are being used to design new polymeric materials, called 'foldamers' [9]. Folded helical bundles have now been designed using non-biological backbones [10••] and there is a rapidly expanding list of such molecules finding applications in biomedicine, including antimicrobials [11], lung surfactant replacements [12], cytomegalovirus inhibitors [13•] and potential siRNA delivery agents [14•].

In addition, protein design is also an increasingly successful enterprise. Novel proteins are now being designed as variants of existing proteins [15•,16•,17], or from broadened alphabets of non-natural amino acids [18] or *de novo*

[19]. However, key challenges remain — to better understand the relative strengths of intermolecular and solvation interactions, and to design a broad range of folds and tighter packing, for example. Nevertheless, questions of great principle are no longer bottlenecks to designing foldable polymers for practical applications and new materials.

Computational protein structure prediction

Bioinformatics based

A long-standing goal of computational biology has been to devise a computer algorithm that takes, as input, an amino acid sequence and gives, as output, the three-dimensional native structure of a protein. A main motivation is to make drug discovery faster and more efficient by replacing slow expensive structural biology experiments with fast cheap computer simulations. A major milestone in computer-based native structure prediction was the invention of CASP (Critical Assessment of Techniques for Structure Prediction) by John Moult, now in its 13th year [20•]. An experiment in the sociology of science, CASP is a community-wide blind test to predict unknown protein structures, given only the amino acid sequence. Currently, homology modeling has the speed to compute approximate folds for large fractions of whole genomes [21,22]. For single-domain globular proteins smaller than about 90 amino acids, web servers can commonly predict native structures often to within about 2–6 Å of their experimental structures [23••,24••,25].

Remaining challenges include predicting the structures of large multidomain or domain-swapped proteins, consistency in achieving errors routinely better than 3 Å and predicting the native states of membrane proteins [26]. Nevertheless, these successes in the computer-based prediction of native protein structures are far beyond what was expected 20 years ago, when the problem looked impossible. The new frontiers are in predicting protein–protein interactions [22,27] and protein function [28].

Physics based

A small scientific community currently aims to use purely physics-based methods, without knowledge derived from databases (such as statistical energy functions or secondary structure predictors), to explore native structures and folding processes. Once 'physics-only' or 'physics-mainly' approaches succeed, the advantages would be: the ability to predict conformational changes, such as induced fit, a common and important unsolved problem in computational drug discovery; the ability to understand protein mechanisms, motions, folding processes, conformational transitions and other situations in which protein behavior requires more than just knowledge of the static native structure; the ability to design synthetic proteins for new applications or to design foldable polymers from non-biological backbones; and the ability to systematically improve protein modeling based on the laws of physics. Physics-based methods are currently limited by some inaccuracies in the force-fields and by huge computational requirements. Nevertheless, there have been notable successes in the past decade. The first milestone was a supercomputer simulation by Duan and Kollman [29] in 1998 of the 36-residue villin headpiece in explicit solvent starting from an unfolded conformation, for nearly a microsecond of computer time, reaching a collapsed state 4.5 Å from the NMR structure. More recently, the IBM Blue Gene group of Pitera and Swope [30] folded the 20-residue Trp-cage peptide in implicit solvent to within ~1 Å using 92 ns of replica-exchange molecular dynamics. With Folding@Home, a distributed grid computing system, Pande *et al.* [31] folded villin to a distance RMSD of 1.7 Å.

There have also been successes in physics-mainly methods, whereby physical potentials are combined with some database information. In summary, although physical models lag behind bioinformatics methods in predicting native structures, the energy functions are proving to be better than thought a few years ago, and distributed computing and new search methods are making inroads into computing large protein conformational changes.

Folding speed and mechanism

In 1968, Cyrus Levinthal first noted the puzzle that, even though proteins have vast conformational spaces, proteins can search and converge quickly to native states, sometimes in microseconds. How do proteins find their native states so quickly? For many of us, this is the fundamental protein folding problem. An interesting conjecture, probably originated by Robert L Baldwin, was that understanding the mechanism of protein folding might lead to fast computational algorithms for predicting native structures from their amino acid sequences.

The question of folding mechanism has driven major advances in folding experiments. Two decades ago, experimentalists had few tools. Key problems included how to measure folding events on timescales faster than a few milliseconds and how to monitor individual chain monomers during folding. Now, we have fast laser temperature-jump methods [32]; mutational methods that give quantities called ϕ [33] or ψ [34] values, which can identify those amino acids that control the folding speed; FRET methods that can watch the formation of particular contacts [35,36]; hydrogen exchange methods that see structural folding events [37]; and extensive studies on model proteins, including cytochrome *c*, chymotrypsin inhibitor 2, barnase, apomyoglobin, src, α-spectrin, fyn SH3 domains, proteins L and G, WW domains, trpzip and the Trp-cage. These studies have led to the recognition, first by Plaxco, Simons and Baker, of a universal property. They found that protein folding

speeds — now known to vary over more than eight orders of magnitude — correlate with the topology of the native protein: fast folders usually have mostly local structure, such as helices and tight turns, whereas slow folders usually have more non-local structure, such as β sheets [38]. Interestingly, however, the fastest known folder at the present time is a three-stranded β sheet; it folds in 140 ns [39••]. A key frontier is to understand the speed limit of folding [40]. Work on 'ultrafast' folders is redefining how we think about kinetic barriers in chemical rate processes [41,42].

Our understanding of folding mechanisms has also been advanced by theory and simulations. Where do we now stand on the matter of the speed principle raised by Levinthal? One step towards an answer was the recognition, through statistical mechanical modeling, that folding does not involve a single microscopic pathway, but rather funnel-shaped energy landscapes [43–45]. The road to the native state from the vast majority of individual non-native conformations is downhill and is different for each non-native starting conformation. Folding processes are microscopically heterogeneous and thus are not readily probed by classical experiments, even despite the advances noted above, because traditional experiments 'see' only average quantities, not variations and distribution functions. Funnels can explain experimental observations that are otherwise paradoxical when interpreted in more classical ways; for example, the finding that transition states would appear to the left of the reactant or to the right of the product when interpreted using simple Hammond–Bronsted models of reaction coordinate diagrams [46–48]. A key unsolved problem remains to rationalize how folding rates change with specific mutations, although a little progress has been made [49]. On the horizon for characterizing kinetic heterogeneity are single-molecule experiments [50•,51••]. Such experiments promise to show us the detailed shapes of folding energy landscapes. Moreover, single-molecule studies might be where experiments meet simulations; master-equation theories are now extending the timescales of physics-based simulations to reach those measured by experiments [52,53••].

However, the folding funnel concept is not a complete answer to Levinthal's puzzle either. The Baldwin conjecture has been a central challenge. To instruct a computer program to find a native state more efficiently than Monte Carlo or molecular dynamics, we need more. We need to know the microscopic folding routes. How does a given chain conformation reach the downhill gulleys that can take it to the native state, and how does it avoid traps and hills? One microscopic mechanism that has been proposed is zipping and assembly (ZA) [54]. According to the ZA mechanism, proteins can fold quickly because they don't search all their degrees of freedom at the same time. Proteins fold over a wide range of timescales. On the fastest timescales (picoseconds to nanoseconds), different small peptide pieces of the chain explore local conformations independently of other such pieces. Local structure forms and then grows (zips) to include increasingly more surrounding chain. Multiple pieces may then assemble together on slower timescales. The key problem this mechanism solves is what conformations a protein does not search [55,56]. Recent tests show that the ZA mechanism speeds up conformational searching sufficiently that physics-only models can now find approximately correct folds for chain lengths up to around 100 monomers [57]. Thus, the ZA mechanism provides a plausible answer to Levinthal's kinetic protein folding problem and shows why proteins don't need supercomputers to guide them to their native structures.

Conclusions

In short, protein folding no longer appears to be an insurmountable grand challenge. Rather, in the words of cartoonist Walt Kelly: ''we now face insurmountable opportunities''. Current knowledge of folding codes is sufficient to guide the successful design of new proteins and new materials. Current computer algorithms are now predicting the native structures of small simple proteins remarkably accurately, contributing to drug discovery and proteomics. Even the once intractable Levinthal puzzle now seems to have a very simple answer: a protein can fold quickly and solve its large global optimization puzzle simply through piecewise solutions of smaller component puzzles.

Update

Two recently published papers are relevant to this review [58•,59•].

Acknowledgements

We thank Kingshuk Ghosh for his assistance and comments. We appreciate the support of National Institutes of Health grant GM 34993 and the Sandler Foundation.

References and recommended reading

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

- • of special interest
- •• of outstanding interest

1. Seringhaus M, Gerstein M: **Chemistry Nobel rich in structure**. *Science* 2007, **315**:40-41.
2. Editorial: **So much more to know**. *Science* 2005, **309**:78-102.
3. Anfinsen CB, Scheraga HA: **Experimental and theoretical aspects of protein folding**. *Adv Protein Chem* 1975, **29**:205-300.
4. Dill KA: **Polymer principles and protein folding**. *Protein Sci* 1999, **8**:1166-1180.
5. Rose GD, Fleming PJ, Banavar JR, Maritan A: **A backbone-based theory of protein folding**. *Proc Natl Acad Sci USA* 2006, **103**:16623-16633.
6. Kamtekar S, Schiffer JM, Xiong H, Babik JM, Hecht MH: **Protein design by binary patterning of polar and nonpolar amino acids**. *Science* 1993, **262**:1680-1685.

7. Schafmeister CE, LaPorte SL, Miercke LJ, Stroud RM: **A designed four helix bundle protein with native-like structure**. *Nat Struct Biol* 1997, **4**:1039-1046.

8. •• Wurth C, Kim W, Hecht MH: **Combinatorial approaches to probe the sequence determinants of protein aggregation and amyloidogenicity**. *Protein Pept Lett* 2006, **13**:279-286.
Using combinatorial methods, Hecht and co-workers show the key sequence features that are related to natural amyloid formation, which helps decode the mechanism of formation of amyloid fibrils.

9. Kirshenbaum K, Zuckermann RN, Dill KA: **Designing polymers that mimic biomolecules**. *Curr Opin Struct Biol* 1999, **9**:530-535.

10. •• Lee BC, Zuckermann RN, Dill KA: **Folding a nonbiological polymer into a compact multihelical structure**. *J Am Chem Soc* 2005, **127**:10999-11009.
Lee *et al*. have made 'foldamers' using a non-biological backbone, in this case, a peptoid chain (*N*-substituted glycine). They show that these chains, up to 60-mers, fold cooperatively into multihelix bundle structures.

11. Lai JR, Huck BR, Weisblum B, Gellman SH: **Design of non-cysteine-containing antimicrobial beta-hairpins: structure-activity relationship studies with linear protegrin-1 analogues**. *Biochemistry* 2002, **41**:12835-12842.

12. Wu CW, Seurynck SL, Lee KY, Barron AE: **Helical peptoid mimics of lung surfactant protein C**. *Chem Biol* 2003, **10**:1057-1063.

13. • English EP, Chumanov RS, Gellman SH, Compton T: **Rational development of beta-peptide inhibitors of human cytomegalovirus entry**. *J Biol Chem* 2006, **281**:2661-2667.
Gellman and colleagues have designed 'β-peptides' that can inhibit human cytomegalovirus (HCMV), a pervasive and significant pathogen, from entering cells, possibly leading to a new class of antiviral agent.

14. • Utku Y, Dehan E, Ouerfelli O, Piano F, Zuckermann RN, Pagano M, Kirshenbaum K: **A peptidomimetic siRNA transfection reagent for highly effective gene silencing**. *Mol Biosyst* 2006, **2**:312-317.
RNAi might be important for the therapeutic silencing of deleterious genes. This research describes the design of a 'lipitoid', a cationic oligopeptoid–phospholipid conjugate, for the non-viral transfection of synthetic siRNA oligos in cell culture, as a possible safe non-toxic delivery vehicle.

15. • Dwyer MA, Looger LL, Hellinga HW: **Computational design of a biologically active enzyme**. *Science* 2004, **304**:1967-1971.
Using structure-based computational guidance, designed mutations have converted ribose-binding protein (lacking enzyme activity) into a glycolytic enzyme, triose phosphate isomerase. The resulting designs showed 10^{-5} to 10^{-6} rate enhancements over the uncatalyzed reaction and are biologically active.

16. • Kaplan J, DeGrado WF: ***De novo* design of catalytic proteins**. *Proc Natl Acad Sci USA* 2004, **101**:11566-11570.
These authors designed a novel catalyst, a phenol oxidase, from first principles. The catalytic peptides were designed by varying the monomer sequences. The peptides self-assembled to bind metal ions at an active site.

17. Lassila JK, Privett HK, Allen BD, Mayo SL: **Combinatorial methods for small-molecule placement in computational enzyme design**. *Proc Natl Acad Sci USA* 2006, **103**:16710-16715.

18. Wang L, Xie J, Schultz PG: **Expanding the genetic code**. *Annu Rev Biophys Biomol Struct* 2006, **35**:225-249.

19. Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL, Baker D: **Design of a novel globular protein fold with atomic-level accuracy**. *Science* 2003, **302**:1364-1368.

20. • Moult J: **Rigorous performance evaluation in protein structure modelling and implications for computational biology**. *Philos Trans R Soc Lond B Biol Sci* 2006, **361**:453-458.
This excellent review summarizes achievements in protein structure prediction methods throughout the six CASP competitions and the current challenges that remain to be solved.

21. Tress M, Ezkurdia I, Grana O, Lopez G, Valencia A: **Assessment of predictions submitted for the CASP6 comparative modeling category**. *Proteins* 2005, **61(suppl 7)**:27-45.

22. Pieper U, Eswar N, Davis FP, Braberg H, Madhusudhan MS, Rossi A, Marti-Renom M, Karchin R, Webb BM, Eramian D *et al.*: **MODBASE: a database of annotated comparative protein structure models and associated resources**. *Nucleic Acids Res* 2006, **34**:D291-D295.

23. •• Bradley P, Misura KM, Baker D: **Toward high-resolution *de novo* structure prediction for small proteins**. *Science* 2005, **309**:1868-1871.
Baker and co-workers show that Rosetta, their very successful bioinformatics-based structure prediction method, has now advanced to the stage of high-resolution structure prediction ($<$1.5 Å) over entire protein structures, for small single-domain proteins.

24. •• Zhang Y, Arakaki AK, Skolnick J: **TASSER: an automated method for the prediction of protein tertiary structures in CASP6**. *Proteins* 2005, **61(suppl 7)**:91-98.
The TASSER method is a highly successful method of computational protein structure prediction based on template identification by threading and 3D structure assembly through the rearrangement of template fragments using database-derived potentials.

25. Baker D: **Prediction and design of macromolecular structures and interactions**. *Philos Trans R Soc Lond B Biol Sci* 2006, **361**:459-463.

26. Bowie JU: **Solving the membrane protein folding problem**. *Nature* 2005, **438**:581-589.

27. Mendez R, Leplae R, Lensink MF, Wodak SJ: **Assessment of CAPRI predictions in rounds 3-5 shows progress in docking procedures**. *Proteins* 2005, **60**:150-169.

28. Pellegrini-Calace M, Soro S, Tramontano A: **Revisiting the prediction of protein function at CASP6**. *FEBS J* 2006, **273**:2977-2983.

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

30. Pitera JW, Swope W: **Understanding folding and design: replica-exchange simulations of ''Trp-cage'' miniproteins**. *Proc Natl Acad Sci USA* 2003, **100**:7587-7592.

31. Zagrovic B, Snow CD, Shirts MR, Pande VS: **Simulation of folding of a small alpha-helical protein in atomistic detail using worldwide-distributed computing**. *J Mol Biol* 2002, **323**:927-937.

32. Yang WY, Gruebele M: **Folding at the speed limit**. *Nature* 2003, **423**:193-197.

33. 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.

34. Sosnick TR, Dothager RS, Krantz BA: **Differences in the folding transition state of ubiquitin indicated by ϕ and ψ analyses**. *Proc Natl Acad Sci USA* 2004, **101**:17377-17382.

35. Magg C, Kubelka J, Holtermann G, Haas E, Schmid FX: **Specificity of the initial collapse in the folding of the cold shock protein**. *J Mol Biol* 2006, **360**:1067-1080.

36. Schuler B, Lipman EA, Eaton WA: **Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy**. *Nature* 2002, **419**:743-747.

37. Maity H, Maity M, Krishna MM, Mayne L, Englander SW: **Protein folding: the stepwise assembly of foldon units**. *Proc Natl Acad Sci USA* 2005, **102**:4741-4746.

38. Plaxco KW, Simons KT, Baker D: **Contact order, transition state placement and the refolding rates of single domain proteins**. *J Mol Biol* 1998, **277**:985-994.

39. •• Xu Y, Purkayastha P, Gai F: **Nanosecond folding dynamics of a three-stranded β-sheet**. *J Am Chem Soc* 2006, **128**:15836-15842.
This study reports the stability and folding kinetics of one of the fastest-folding proteins: a three-stranded β sheet. It folds in around 140 ns at 35.0 °C.

40. Kubelka J, Hofrichter J, Eaton WA: **The protein folding 'speed limit'**. *Curr Opin Struct Biol* 2004, **14**:76-88.

41. Sadqi M, Fushman D, Munoz V: **Atom-by-atom analysis of global downhill protein folding**. *Nature* 2006, **442**:317-321.

42. Huang F, Sato S, Sharpe TD, Ying L, Fersht AR: **Distinguishing between cooperative and unimodal downhill protein folding**. *Proc Natl Acad Sci USA* 2007, **104**:123-127.

43. Dill KA: **Theory for the folding and stability of globular proteins**. *Biochemistry* 1985, **24**:1501-1509.

44. Leopold PE, Montal M, Onuchic JN: **Protein folding funnels: a kinetic approach to the sequence-structure relationship**. *Proc Natl Acad Sci USA* 1992, **89**:8721-8725.

45. Dill KA, Chan HS: **From Levinthal to pathways to funnels**. *Nat Struct Biol* 1997, **4**:10-19.

46. Wright CF, Lindorff-Larsen K, Randles LG, Clarke J: **Parallel protein-unfolding pathways revealed and mapped**. *Nat Struct Biol* 2003, **10**:658-662.

47. Levy Y, Cho SS, Onuchic JN, Wolynes PG: **A survey of flexible protein binding mechanisms and their transition states using native topology based energy landscapes**. *J Mol Biol* 2005, **346**:1121-1145.

48. Ozkan SB, Bahar I, Dill KA: **Transition states and the meaning of ϕ-values in protein folding kinetics**. *Nat Struct Biol* 2001, **8**:765-769.

49. Weikl TR, Dill KA: **Transition-states in protein folding kinetics: the structural interpretation of ϕ values**. *J Mol Biol* 2007, **365**:1578-1586.

50. • Sherman E, Haran G: **Coil-globule transition in the denatured state of a small protein**. *Proc Natl Acad Sci USA* 2006, **103**:11539-11543.
This single-molecule study characterizes the equilibrium coil-globule transition in the denatured state of the Ig-binding domain of protein L. Using FRET measurements of individual protein molecules, the radius of gyration of the protein at different denaturant concentration is determined. Analysis reveals that the fully collapsed state of the protein is only $\sim$10% larger than the folded state.

51. •• Cecconi C, Shank EA, Bustamante C, Marqusee S: **Direct observation of the three-state folding of a single protein molecule**. *Science* 2005, **309**:2057-2060.
This work studies the unfolding and refolding of single ribonuclease H (RNase H) protein molecules, using optical tweezers. The study provides a direct single-molecule observation of an intermediate state.

52. Chodera JD, Swope WC, Pitera JE, Dill KA: **Long-time protein folding dynamics from short-time molecular dynamics simulations**. *Multiscale Model Simul* 2006, **54**:1214-1226.

53. •• Jayachandran G, Vishal V, Pande VS: **Using massively parallel simulation and Markovian models to study protein folding: examining the dynamics of the villin headpiece**. *J Chem Phys* 2006, **124**:164902.
This is a massive computational study of a small single-domain protein, the villin headpiece. Using large-scale distributed computing, all-atom explicit solvent molecular dynamics (MD) simulations are sampled up to 500 μs in the form of ten of thousands of independent MD trajectories. Long-timescale kinetics are then obtained by constructing Markov models from the MD trajectories.

54. Dill KA, Fiebig KM, Chan HS: **Cooperativity in protein-folding kinetics**. *Proc Natl Acad Sci USA* 1993, **90**:1942-1946.

55. Hockenmaier J, Joshi AK, Dill KA: **Routes are trees: the parsing perspective on protein folding**. *Proteins* 2007, **66**:1-15.

56. Voelz VA, Dill KA: **Exploring zipping and assembly as a protein folding principle**. *Proteins* 2007, **66**:877-888.

57. Ozkan SB, Wu GA, Chodera JD, Dill KA: **Protein folding by zipping and assembly**. *Proc Natl Acad Sci USA* 2007, in press.

58. • Mukhopadhyay S, Krishnan R, Lemke EA, Lindquist S, Deniz AA: **A natively unfolded yeast prion monomer adopts an ensemble of compact and fluctuating structures**. *Proc Natl Acad Sci USA* 2007, **104**:2649-2654.
A single-molecule FRET study of a yeast prion protein, giving the full folding time distribution function.

59. • Lei H, Wu C, Liu H, Duan Y: **Folding free energy landscape of villin headpiece subdomain from molecular dynamic simulations**. *Proc Natl Acad Sci USA* 2007, **104**:4925-4930.
These authors show accurate folding of the 36-mer villin headpiece protein using a physical all-atom model with implicit solvent.