CURRENT OPINION
www.current-opinion.com

Protein folding: bringing theory and experiment closer together

Michele Vendruscolo* and Emanuele Paci†

The ability to perform enzymatic function by harnessing random molecular motion into self-organized protein structures is one of the most fascinating results of evolution. A close interplay between theory and experiment is driving the progress in understanding the principles that determine the behaviour of proteins. New techniques that significantly increase the amount of information obtainable from experimental data have been recently proposed; it is now becoming possible to describe at atomic resolution the events that take place during the folding process. Successful predictions of these events are being reported at an increasing rate and general principles are being outlined.

Addresses
*Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom
e-mail: mv245@cam.ac.uk
†Biochemisches Institut der Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland
e-mail: paci@bioc.unizh.ch

Current Opinion in Structural Biology 2003, **13**:82–87

This review comes from a themed issue on
Folding and binding
Edited by Jane Clarke and Gideon Schreiber

0959-440X/03/$ – see front matter

DOI 10.1016/S0959-440X(03)00007-1

Introduction

The concept of the free energy landscape has promoted much of the recent progress in understanding the process of protein folding [1,2••,3,4••]. Two essential principles of protein folding are summarized by this picture. The first is that folding is a stochastic process in which the free energy decreases spontaneously and the second is that evolution has selected amino acid sequences that avoid misfolding, long-lived metastable traps and aggregation. Equally important is the fact that landscapes provide a convenient framework for visualizing and interpreting experimental data on the thermodynamics and the kinetics of proteins.

The common goal for theoreticians and experimentalists is to determine the folding process at atomic resolution by combining experimental and computational results. Crucial to this end are recent advances in experimental and in computational techniques that allow the processes at the individual residue level to be reconstructed [2••,4••]. Two complementary approaches have been gradually emerging and they now play a major role in the analysis of the vast amount of experimental knowledge accumulated about the process of protein folding. On one hand, methods to extract more information from experimental data have been proposed [5•,6•,7]. On the other hand, such methods are complemented by theoretical calculations that have been able to reproduce experimental results and, in some cases, predict new ones [1,2••,4••,8].

Experimental techniques

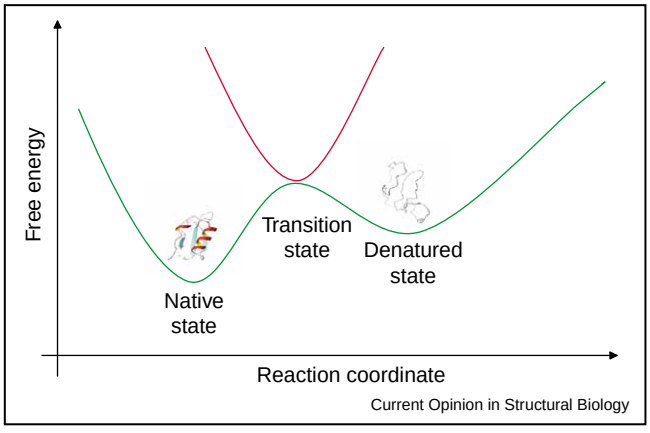
Advances in experimental techniques have made it possible to obtain detailed information about the successive conformations that occur during the folding process [2••,4••,9–11]. Protein engineering is providing residue-specific information about the structures of intermediates and transition states for folding [4••]. Several NMR [4••,6•,12,13•,14–16], hydrogen exchange [2••,10], fluorescence transfer [17,18] and triplet excitation [9] techniques have been developed to detect long-range interactions in partially folded states and in unfolded states, thus revealing how native order originates from the very first stages of the folding process. By using atomic force microscopy, proteins are unfolded by applying a mechanical force [19,20]. With this technique, experiments are performed under non-equilibrium conditions and a reaction coordinate (the end-to-end distance) is imposed on the unfolding process. The effect of mutations on the measured forces makes it possible to probe the pathways of forced unfolding [19,21]. A method for quantitatively probing forced unfolding pathways has been recently proposed by Jane Clarke's group [22•,23]. This method extends the technique of ϕ-value analysis [4••] to provide a high-resolution picture of the transition state for forced unfolding.

Parallel to these developments, advances in computational methods have made it possible to significantly increase our ability to translate experimental data into structural information about the states that are populated during the folding process. An ensemble of conformations corresponding to the unfolded state of the drkN SH3 domain protein was obtained by a procedure in which structures generated by a fast simulation technique that are most compatible with a series of NMR data were selected [24]. A similar filtering procedure was shown to be effective in determining ensembles of conformations that are compatible with small-angle X-ray scattering (SAXS) data, which give information about the pairwise distance distribution function of a protein in solution [7]. As more SAXS data on folding are becoming available, this technique is promising for the large-scale structural determination needed in structural genomics. A related method of selection, in which structures generated by molecular dynamics unfolding trajectories are tested for

compatibility with protein engineering data (ϕ values), was used recently to determine and validate the transition state for folding of the src SH3 domain protein [25•].

In a recent development, it has been shown that the amount of information that can be obtained from experimental measurements can be expanded further by using the data to build up phenomenological energy functions to bias computer-generated trajectories. With this approach, conformations compatible with experimental data are determined directly during the simulations [5•,26], rather than being obtained from filtering procedures, such as those discussed above [7,24,25•]. Experimental data are incorporated into the energy function to transform the landscape and create a minimum in correspondence of the state observed experimentally and therefore allows very efficient sampling of conformational space (see Figure 1). The transition state for folding of acylphosphatase, a 98-residue two-state protein, was determined in this way [5•,26] (see Figure 2). These studies allowed new insight to be obtained into the determinants of the folding process by showing that the network of interactions that stabilizes the transition state is established when a few key residues form their native-like arrangement [5•,26,27]. Using an approach of this type, it has also been possible to determine the experimental free energy landscape corresponding to the molten globule state of α-lactalbumin; NMR data at increasing concentrations of denaturant were used to probe different regions of the conformational space (M Vendruscolo, E Paci, M Karplus, CM Dobson, unpublished data). This analysis reveals that free energy landscapes of proteins are characterized by deep valleys that are robust against changes in the external conditions.

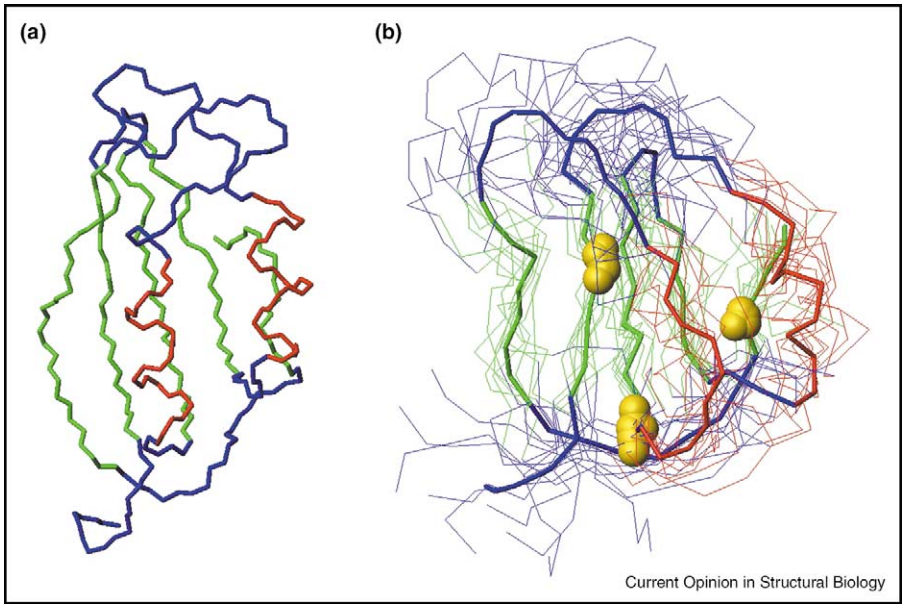
Figure 1



Schematic representation of the transformation of the energy landscape used to determine the structure of the transition state for folding from experimental ϕ-value data [5•,26]. The transition state is a saddle point in the actual landscape (green line), whereas it is the global minimum of the transformed landscape (red line).

The combined use of experimental data and computer simulations has also been shown to be very effective in structural genomics for protein structure determination [28]. In this case, high-throughput residual dipolar

Figure 2



Comparison of **(a)** the native state structure and **(b)** the most representative structures of the transition state ensemble of acylphosphatase, determined by all-atom molecular dynamics simulations [26]. Native secondary structure elements are shown in colour (the two α helices are plotted in red and the β sheet is in green). The three key residues for folding are shown as gold spheres [5•,26].

couplings were used to build up a scoring function to guide the generation of protein structures from a stochastic assembly of small fragments.

Computational techniques

Advances in computer technology and algorithms have made it possible to perform simulations at atomic-level accuracy, thus narrowing the gap between theory and experiment [4••,29–31].

Molecular dynamics has been especially successful in generating unfolding trajectories [2••,4••]. In several cases, detailed predictions from molecular dynamics simulations have been confirmed experimentally [4••]. Furthermore, growing evidence indicates that unfolding trajectories, near the native state in particular, are close to the folding trajectories [2••,4••]. An explanation for the origin of this reversibility is that molecular evolution has created, as mentioned above, deep valleys in the free energy landscape that confer robustness against perturbations, such as an increase in temperature or chemical denaturation (M Vendruscolo, E Paci, M Karplus, CM Dobson, unpublished data). As unfolding trajectories, even at high temperature, are restricted to these valleys, they are bound to follow paths similar to those of the folding trajectories at room temperature. The existence of deep valleys also offers a natural way to define reaction coordinates for the folding process, because all the upward movements from the bottom of the valley are quickly averaged out. We therefore suggest that a 'principle of mesoscopic reversibility' should hold for the process of protein folding. This principle is a direct consequence of the evolutionary pressure against misfolding and aggregation. Therefore, it is the negative design of non-native states that has created the deep valleys that make free energy landscapes robust against perturbations. The principle of mesoscopic reversibility should also hold for other evolved systems of macromolecular size. Deep valleys have been reported for protein–protein binding [32] and for RNA folding [33]. It has also been suggested that the existence of one or very few pathways acts as an evolved quality control mechanism to avoid misfolding [34].

The most relevant development in the field of computational studies is the possibility of generating long folding trajectories using molecular dynamics simulations based on accurate force-fields. The first cases reported involved small α proteins [31] and β peptides [30,35]. Small proteins with extremely fast folding times, such as the engrailed homeodomain [36] and the WW domain [37], are suitable candidates for future studies.

To provide a structural interpretation of the typical sawtooth-like spectra measured in single-molecule stretching experiments, various simulation techniques have been proposed, differing mainly in the way the solvent is treated, in which detailed all-atom models of proteins are stretched by pulling two atoms apart [38,39]. As recently reviewed in this journal [40], in some cases, simulations can effectively explain the force patterns measured. It should be borne in mind, however, that the forced unfolding of proteins is a nonequilibrium phenomenon strongly dependent on the pulling speed and, because the timescales of simulations and experiments are very different, the respective pathways need not to be the same. Recently, through a combination of experimental analysis and molecular dynamics simulations, it has been shown that, in the case of mechanical unfolding, pathways might effectively be the same for a large range of pulling speeds or forces [22•,23], thus providing another demonstration of the robustness of the free energy landscapes.

A challenge for computer simulations is that a complete description of the folding process involves the determination of a large number of trajectories, because folding is a stochastic process characterized by significant heterogeneity. The presence of broad distribution functions is illustrated clearly by single-molecule experiments. For example, it was found that the folding time of individual titin molecules, as measured by fluorescence quenching, can vary at least from 10 to 70 ms [41]. Such heterogeneity has prompted a method in which a massive number of short folding trajectories are generated by computer simulations to access the leftmost tail of the distribution of folding times [29]. With this method, it is possible to characterize the folding of proteins with lag times not longer than the length of molecular dynamics simulations feasible on a personal computer (tens of nanoseconds) and for which the fastest folding trajectories coincide with the typical ones [42].

Simple theoretical models

In the past decade, theoretical approaches based on simplifying assumptions have been contributing to the clarification of the principles of protein folding [1,2••,8]. The most important aspect of simple models is that they enable the complete folding process to be followed. Therefore, the events described by these simulations suggest generic mechanisms that may operate in proteins. The nucleation mechanism for protein folding, for example, can be revealed by simulating protein-like polymers on a lattice [2••,43]. Lattice simulations have also been recently used to investigate the origin of cooperativity in protein folding [44] and the effect of a pulling force on protein folding [45]. Recent results from a coarse-grained model of protein folding suggest that the events during the initial collapse are independent of the amino acid sequence, although naturally occurring sequences show an increased tendency to form specific native-like interactions [46]. Simple models based on the explicit use of the native topology have also been used to study the equilibrium and kinetic properties of a protein in the

presence of a pulling force [47–49]. The results are consistent with those obtained with more detailed models and under strong nonequilibrium conditions, suggesting that the process of forced unfolding may be dominated by steric interactions and that the energy landscape is robust against variations in magnitude of the external force.

Predictions from simple models should be carefully tested experimentally [2••,4••,9]. An interesting debate was sparked by the suggestion that the kinetics of folding is determined by the difference between the temperatures of collapse and of folding [50]. Later experiments, however, failed to find proteins whose folding kinetics were well described by such a prediction [51]. Another prediction that awaits experimental confirmation is that nonclassical ϕ values originate from the presence of multiple folding pathways [52]. In a recent study [53], the folding process of protein G was studied in all-atom Monte Carlo simulations with a force-field in which the native interactions were assigned favourable contributions (the so-called Go model) [53,54]. Single-exponential fluorescence data were shown to be compatible with the presence of multiple folding pathways [53]. Although the folding mechanism in a Go model depends on the relative weight given to the native interactions [54,55], these simulations suggest a possible nontrivial process behind the single-exponential behaviour measured from ensemble averages and require that further experiments be carried out. Growing evidence also shows that simple models based on the knowledge of the native topology can identify correct features of the mechanism of folding for two-state proteins in a significant number of cases [27,55,56]. It has also been shown that the thermodynamic analysis of a Go-like model was capable of identifying sites in HIV-1 protease that play a key role in causing resistance to protease-inhibiting drugs [57].

Conclusions and future prospects

The landscape picture has provided a powerful conceptual framework to rationalize experimental observations and to suggest new measurements. At the same time, advances in simulation techniques and computer technology have allowed an atomistic description of the folding process, thus creating the possibility of obtaining a direct representation of the events that take place during folding. Parallel to these theoretical developments, advances in experimental techniques, now capable of providing residue-specific information about the dynamics of folding, have created a situation in which details of the folding process can actually be detected with exquisite precision. These results represent a further challenge for theoreticians to refine the conceptual framework to include new observations.

The progress in understanding protein folding is already opening the way toward solving more complex problems, such as those concerning the mechanics of misfolding and aggregation [58]. At present, the lack of a precise knowledge of the structure of the aggregates is a tremendous disadvantage with respect to protein folding. The generalization of experimental and computational techniques developed for the study of protein folding is suggesting effective ways to gain insight into the structure of fibrillar aggregates and the process of their formation.

The possibility of computing the folding process at the atomic level using molecular dynamics is a remarkable achievement that will gradually enable us to understand the fundamental processes involved in protein folding. Such an understanding will possibly suggest new, more coarse-grained models capable of capturing the principles of biomolecular self-organization. Alternatively, it may be that all-atom models with detailed physicochemical energy functions cannot be further simplified. This, however, will be a surprising finding because we are accustomed to the existence of macroscopic laws, as for example the rate equations in chemical kinetics, that ultimately describe phenomenologically the emergence of some kind of regular behaviour when a very large number of interacting elements are considered [59]. One possible problem is that random fluctuations caused by thermal motion may undermine any regularity at the mesoscopic level of macromolecules [60]. A possible role of evolution may have been to reduce the impact of thermal fluctuations by selecting systems robust against perturbations. Among evolved systems, proteins are likely to be an ideal candidate for discovering the organizing principles at the molecular level — increasing evidence already supports the existence of regularities, such as the principle of minimal frustration [1] and the principle of mesoscopic reversibility.

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. Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG: **Funnels, pathways, and the energy landscape of protein folding: a synthesis**. *Proteins* 1995, **21**:167-195.

2. •• Dinner AR, Sali A, Smith LJ, Dobson CM, Karplus M: **Understanding protein folding via free-energy surfaces from theory and experiment**. *Trends Biochem Sci* 2000, **25**:331-339.

This comprehensive review presents a summary of recent experimental and theoretical advances.

3. Brooks CL, Onuchic JN, Wales DJ: **Statistical thermodynamics. Taking a walk on a landscape**. *Science* 2001, **293**:612-613.

4. •• Fersht AR, Daggett V: **Protein folding and unfolding at atomic resolution**. *Cell* 2002, **108**:573-582.

This review discusses progress in closing the gap between theory and experiment.

5. • Vendruscolo M, Paci E, Dobson CM, Karplus M: **Three key residues form a critical contact network in a protein folding transition state**. *Nature* 2001, **409**:641-645.

This paper introduces an innovative approach in which experimental data are used as restraints for the determination of the structure of the transition state for folding.

6. • Choy WY, Mulder FA, Crowhurst KA, Muhandiram DR, Millett IS, Doniach S, Forman-Kay JD, Kay LE: **Distribution of molecular size within an unfolded state ensemble using small-angle X-ray scattering and pulse field gradient NMR techniques**. *J Mol Biol* 2002, **316**:101-112.

An effective approach, based on a combination of experimental and computational techniques, is used to analyse the conformational space of a protein whose native state exists in equilibrium with the unfolded state.

7. Zheng W, Doniach S: **Protein structure prediction constrained by solution X-ray scattering data and structural homology identification**. *J Mol Biol* 2002, **316**:173-187.

8. Mirny L, Shakhnovich E: **Protein folding theory: from lattice to all-atom models**. *Annu Rev Biophys Biomol Struct* 2001, **30**:361-396.

9. Eaton WA, Muñoz V, Hagen SJ, Jas GS, Lapidus LJ, Henry ER, Hofrichter J: **Fast kinetics and mechanisms in protein folding**. *Annu Rev Biophys Biomol Struct* 2000, **29**:327-359.

10. Rumbley J, Hoang L, Mayne L, Englander SW: **An amino acid code for protein folding**. *Proc Natl Acad Sci USA* 2001, **98**:105-112.

11. Myers JK, Oas TG: **Mechanism of fast protein folding**. *Annu Rev Biochem* 2002, **71**:783-815.

12. Yi Q, Scalley-Kim ML, Alm EJ, Baker D: **NMR characterization of residual structure in the denatured state of protein L**. *J Mol Biol* 2000, **299**:1341-1351.

13. • Kortemme T, Kelly MJ, Kay LE, Forman-Kay J, Serrano L: **Similarities between the spectrin SH3 domain denatured state and its folding transition state**. *J Mol Biol* 2000, **297**:1217-1229.

This NMR study shows that native interactions are also present in the denatured state of a small globular protein.

14. Klein-Seetharaman J, Oikawa M, Grimshaw SB, Wirmer J, Duchardt E, Ueda T, Imoto T, Smith LJ, Dobson CM, Schwalbe H: **Long-range interactions within a nonnative proteins**. *Science* 2002, **295**:1719-1722.

15. McParland VJ, Kalverda AP, Homans SW, Radford SE: **Structural properties of an amyloid precursor of β(2)-microglobulin**. *Nat Struct Biol* 2002, **9**:326-331.

16. Lietzow M, Jamin M, Dyson JH, Wright P: **Mapping long-range contacts in a highly unfolded protein**. *J Mol Biol* 2002, **322**:655-662.

17. Teilum K, Maki K, Kragelund BB, Poulsen FM, Roder H: **Early kinetic intermediate in the folding of acyl-CoA binding protein detected by fluorescence labeling and ultrarapid mixing**. *Proc Natl Acad Sci USA* 2002, **99**:9807-9812.

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

19. Fisher TE, Marszalek PE, Fernandez JM: **Stretching single molecules into novel conformations using the atomic force microscope**. *Nat Struct Biol* 2000, **7**:719-724.

20. Zhuang X, Rief M: **Single-molecule folding.** *Curr Opin Struct Biol* 2003, **13**:in press.

21. Marszalek PE, Lu H, Li H, Carrion-Vazquez M, Oberhauser AF, Schulten K, Fernandez JM: **Mechanical unfolding intermediates in titin modules**. *Nature* 1999, **402**:100-103.

22. • Best RB, Fowler SB, Toca-Herrera JL, Clarke J: **A simple method for probing the mechanical unfolding pathway of proteins in detail**. *Proc Natl Acad Sci USA* 2002, **99**:12143-12148.

The authors present a quantitative approach for analysing atomic force microscopy results for single-molecule unfolding and an efficient way of measuring ϕ values for probing the transition state for mechanical unfolding.

23. Fowler S, Best R, Toca-Herrera JL, Rutherford T, Steward A, Paci E, Karplus M, Clarke J: **Mechanical unfolding of a titin Ig domain: structure of unfolding intermediate revealed by combining AFM, molecular dynamics simulations, NMR and protein engineering**. *J Mol Biol* 2002, **322**:841-849.

24. Choy WY, Forman-Kay JD: **Calculation of ensembles of structures representing the unfolded state of an SH3 domain**. *J Mol Biol* 2001, **308**:1011-1032.

25. • Gsponer J, Caflisch A: **Molecular dynamics simulations of protein folding from the transition state**. *Proc Natl Acad Sci USA* 2002, **99**:6719-6724.

The transition state for folding of an SH3 domain protein is determined and validated by direct refolding using molecular dynamics simulations.

26. Paci E, Vendruscolo M, Dobson CM, Karplus M: **Determination of a transition state at atomic resolution from protein engineering data**. *J Mol Biol* 2002, **324**:151-163.

27. Vendruscolo M, Dokholyan NV, Paci E, Karplus M: **A small-world view of the amino acids that play a role in protein folding**. *Phys Rev E* 2002, **65**:061910.

28. Rohl CA, Baker D: ***De novo* determination of protein backbone structure from residual dipolar couplings using Rosetta**. *J Am Chem Soc* 2002, **124**:2723-2729.

29. Snow CD, Nguyen H, Pande VS, Gruebele M: **Absolute comparison of simulated and experimental protein-folding dynamics**. *Nature* 2002, **420**:102-106.

30. Cavalli A, Ferrara P, Caflisch A: **Weak temperature dependence of the free energy surface and folding pathways of structured peptides**. *Proteins* 2002, **47**:305-314.

31. Ghosh A, Elber R, Scheraga HA: **An atomically detailed study of the folding pathways of protein A with the stochastic difference equation**. *Proc Natl Acad Sci USA* 2002, **99**:10394-10398.

32. Schreiber G: **Kinetic studies of protein-protein interactions**. *Curr Opin Struct Biol* 2002, **12**:41-47.

33. Russell R, Zhuang X, Babcock HP, Millett IS, Doniach S, Chu S, Herschlag D: **Exploring the folding landscape of a structured RNA**. *Proc Natl Acad Sci USA* 2002, **99**:155-160.

34. Davis R, Dobson CM, Vendruscolo M: **Determination of the structures of distinct transition state ensembles for a β-sheet peptide with parallel folding pathways**. *J Chem Phys* 2002, **117**:9510-9517.

35. Smith LJ, Daura X, van Gunsteren WF: **Assessing equilibration and convergence in biomolecular simulations**. *Proteins* 2002, **48**:487-496.

36. Mayor U, Johnson CM, Daggett V, Fersht AR: **Protein folding and unfolding in microseconds to nanoseconds by experiment and simulation**. *Proc Natl Acad Sci USA* 2000, **97**:13518-13522.

37. Ferguson N, Johnson CM, Macias M, Oschkinat H, Fersht A: **Ultrafast folding of WW domains without structured aromatic clusters in the denatured state**. *Proc Natl Acad Sci USA* 2001, **98**:13002-13007.

38. Lu H, Isralewitz B, Krammer A, Vogel V, Schulten K: **Unfolding of titin immunoglobulin domains by steered molecular dynamics simulation**. *Biophys J* 1998, **75**:662-671.

39. Paci E, Karplus M: **Forced unfolding of fibronectin type 3 modules: an analysis by biased molecular dynamics simulations**. *J Mol Biol* 1999, **288**:441-459.

40. Isralewitz B, Gao M, Schulten K: **Steered molecular dynamics and mechanical functions of proteins**. *Curr Opin Struct Biol* 2001, **11**:224-230.

41. Zhuang X, Ha T, Kim HD, Centner T, Labeit S, Chu S: **Fluorescence quenching: a tool for single-molecule protein-folding study**. *Proc Natl Acad Sci USA* 2000, **97**:14241-14244.

42. Fersht AR: **On the simulation of protein folding by short time scale molecular dynamics and distributed computing**. *Proc Natl Acad Sci USA* 2002, **99**:14122-14125.

43. Shakhnovich E, Abkevich V, Ptitsyn O: **Conserved residues and the mechanism of protein folding**. *Nature* 1996, **379**:96-98.

44. Kaya H, Chan HS: **Towards a consistent modeling of protein thermodynamic and kinetic cooperativity: how applicable is the transition state picture to folding and unfolding?** *J Mol Biol* 2002, **315**:899-909.

45. Socci ND, Onuchic JN, Wolynes PG: **Stretching lattice models of protein folding**. *Proc Natl Acad Sci USA* 1999, **96**:2031-2035.

46. Sosnick TR, Berry RS, Colubri A, Fernandez A: **Distinguishing foldable proteins from nonfolders: when and how do they differ?** *Proteins* 2002, **49**:15-23.

47. Klimov DK, Thirumalai D: **Native topology determines force-induced unfolding pathways in globular proteins**. *Proc Natl Acad Sci USA* 2000, **97**:7254-7259.

48. Shen T, Canino LS, McCammon JA: **Unfolding proteins under external forces: a solvable model under the self-consistent pair contact probability approximation**. *Phys Rev Lett* 2002, **89**:068103.

49. Cieplak M, Hoang TX, Robbins MO: **Thermal folding and mechanical unfolding pathways of protein secondary structures**. *Proteins* 2002, **49**:104-113.

50. Thirumalai D, Klimov DK: **Deciphering the timescales and mechanisms of protein folding using minimal off-lattice models**. *Curr Opin Struct Biol* 1999, **9**:197-207.

51. Millet IS, Townsley LE, Chiti F, Doniach S, Plaxco KW: **Equilibrium collapse and the kinetic 'foldability' of proteins**. *Biochemistry* 2002, **41**:321-325.

52. 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.

53. Shimada J, Shakhnovich EI: **The ensemble folding kinetics of protein G from an all-atom Monte Carlo simulation**. *Proc Natl Acad Sci USA* 2002, **99**:11175-11180.

54. Zhou Y, Karplus M: **Interpreting the folding kinetics of helical proteins**. *Nature* 1999, **401**:400-403.

55. Alm E, Morozov AV, Kortemme T, Baker D: **Simple physical models connect theory and experiment in protein folding kinetics**. *J Mol Biol* 2002, **322**:463-476.

56. Guerois R, Serrano L: **The SH3-fold family: experimental evidence and prediction of variations in the folding pathways**. *J Mol Biol* 2000, **304**:967-982.

57. Micheletti C, Cecconi F, Flammini A, Maritan A: **Crucial stages of protein folding through a solvable model: predicting target sites for enzyme-inhibiting drugs**. *Prot Sci* 2002, **11**:1878-1887.

58. Dobson CM: **Getting out of shape**. *Nature* 2002, **418**:729-730.

59. Anderson PW: **More is different**. *Science* 1972, **177**:393-396.

60. Laughlin RB, Pines D, Schmalian J, Stojkovic BP, Wolynes P: **The middle way**. *Proc Natl Acad Sci USA* 2000, **97**:32-37.