ELSEVIER

Theory of protein folding

José Nelson Onuchic[1,2,*] and Peter G Wolynes[1,2,3]

Protein folding should be complex. Proteins organize themselves into specific three-dimensional structures, through a myriad of conformational changes. The classical view of protein folding describes this process as a nearly sequential series of discrete intermediates. In contrast, the energy landscape theory of folding considers folding as the progressive organization of an ensemble of partially folded structures through which the protein passes on its way to the natively folded structure. As a result of evolution, proteins have a rugged funnel-like landscape biased toward the native structure. Connecting theory and simulations of minimalist models with experiments has completely revolutionized our understanding of the underlying mechanisms that control protein folding.

Addresses
[1]Center for Theoretical Biological Physics, [2]Department of Physics, [3]Department of Chemistry and Biochemistry, University of California at San Diego, 9500 Gilman Drive, La Jolla, California 92093, USA
*e-mail: jonuchic@ucsd.edu

Current Opinion in Structural Biology 2004, **14**:70–75

This review comes from a themed issue on
Folding and binding
Edited by David Baker and William A Eaton

0959-440X/$ – see front matter

DOI 10.1016/j.sbi.2004.01.009

Introduction

Protein folding should be complex. Proteins organize themselves into specific three-dimensional structures, through a myriad of conformational changes. Each conformational change is itself a complex solvent-influenced event. So, in detail, a folding mechanism must involve a complex network of elementary reactions. However, simple empirical patterns of protein folding kinetics, such as linear free energy relationships, have been shown to exist.

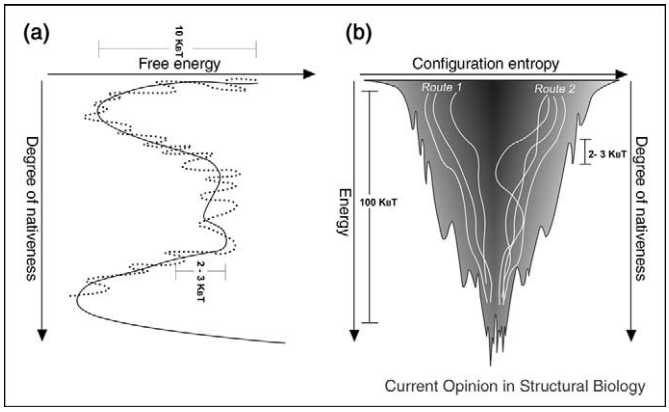
This simplicity is owed to the global organization of the landscape of the energies of protein conformations into a funnel (Figure 1). This organization is not characteristic of all polymers with any sequence of amino acids, but is a result of evolution. The discovery of simple kinetic patterns [1–5] and the existence of a theoretical framework based on the global properties of the energy landscape [6,7] have, in recent years, allowed a very fruitful collaboration between theory and experiment in the study of folding. This review focuses on recent achievements of this collaboration. We wish to highlight what can be learned from the simplest models and leave others to review the results of highly detailed all-atom simulation.

Basic concepts

The locations of atoms in proteins can be determined, in favorable cases, to an accuracy of less than 3 Å using X-ray crystallography. This specificity of structure arises from the heterogeneity of the protein chain. The differing energies associated with positioning different residues near or far from each other or from solvent enable some structures to be more stable than others. If a sequence is chosen at random, the specificity of structure is still small — a variety of globally different structures have very low energies, but within a few k_BT of each other. This mathematical observation has been the bane of computational protein structure prediction using energy functions. The native sequence appears random to a poor energy function and this allows many very different structures to be candidate minimum energy predictions [8]. In the laboratory, if such low-energy structures were to exist for proteins, as they indeed do for random sequences, they would act as kinetic traps, leading to kinetic complexity with very many intermediates. A typical heteropolymer chosen at random has a 'rugged' energy landscape and the ensuing dynamics are 'glassy' [9]. The exact ground state of such a sequence arises by chance from competition between many conflicting energy contributions. The various interactions are 'frustrated'. Even a minor change in the sequence (a single mutation) would usually cause a structurally disparate kinetic trap to become the new ground state. This would be unbeneficial to an evolving line of organisms. If a particular random heteropolymer had already achieved a functionally useful three-dimensional shape in an ancestral organism, a single mutation would usually destroy this useful shape, so the new mutated organism would die. Natural selection over time would not keep such outcomes because of the biological functional constraints [1]. Evolution achieves robustness by selecting for sequences in which the interactions present in the functionally useful structure are not in conflict, as in a random heteropolymer, but instead are mutually supportive and cooperatively lead to a low-energy structure. The interactions are 'minimally frustrated' [6] or 'consistent' [10].

The notion of minimal frustration is not a set of vague words. It has been made quantitatively precise using the statistical mechanics of spin glasses [1,2,9]. The degree of frustration is measured by comparing a folding transition temperature (T_f) with a glass transition temperature (T_g), which characterizes the thermodynamics of trapping. We can think of T_f as characterizing the native interactions

Figure 1

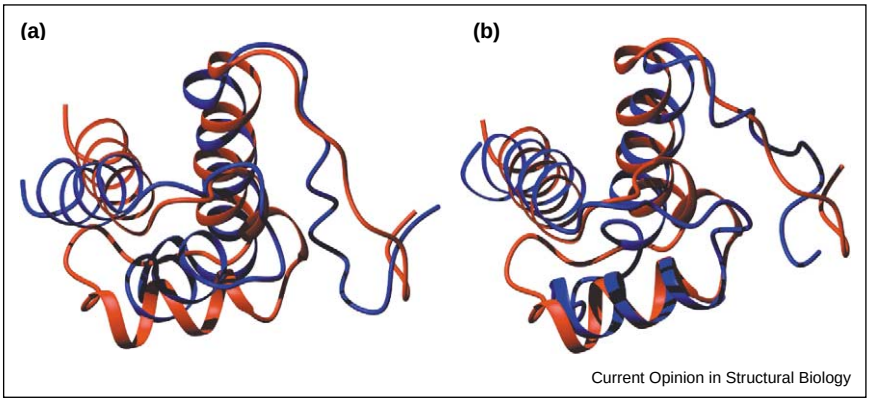


Evolution works on a rather crude energy scale; on this scale, folding landscapes are robust and funnel like **(b)**. Disallowed mutations may cost as much as 10 k_BT. The choice of detailed folding mechanism occurs at a much finer energy scale, a little larger than the thermal fluctuations (reflected in the schematic representation of the traps in the funnel that are a few k_BT deep). Barriers to folding [around k_BT **(a)**] are a relatively small fraction of the total binding energy (around 100 k_BT) of the native folded structure, because a good deal of the binding must be cancelled by a concomitant entropy loss (this is the renormalized energy including effects of the solvation free energy). Folding and unfolding barriers that are the result of the non-perfect cancellation between energy and configurational entropy are subtler still in energy scale, but can exceed k_BT. Thus, an experiment may seem to reveal dramatically different mechanisms for two very similar sequences that fold to the same structure (represented by the ensemble of paths called route 1 or route 2). However, the overall funnel-like landscape is basically the same in both cases (figure prepared by Jason Wiskerchen).

found in the folded structure and T_g as measuring the strength of the non-native interactions that would be present in other configurations. Simulations of simple models show that sequences with a high T_f/T_g ratio fold faster than most sequences [11,12] and with few intermediates. Although older energy functions for structure prediction lead to significant frustration, the quantitative minimal frustration principle has been successfully used to dramatically improve energy functions for structure prediction, as shown in Figure 2 [13,14]. The quantitative theory of minimal frustration also provides an automated algorithm for designing sequences that can efficiently fold to nearly unique structures [15]. There is much evidence that real sequences have evolved to have less frustration than even our best existing energy functions predict.

For an evolved protein, the mutually supportive, cooperative and consistent nature of the interactions in the

Figure 2



Results obtained using an energy function whose parameters are learned from the variational form of the minimal frustration (MFP). Structure prediction results for an FF domain (69 residues) from human HYPA/FBP11 (PDB code 1H40; CASP5 target T0170). **(a)** Superposition of the best structure submitted to CASP5 (blue), from the Schulten-Wolynes groups, and the native structure (red) (Cα rmsd 5.6 Å). **(b)** Superposition of the best Q-score predicted structure from a newly developed associative memory (AM)/water potential (blue) and the native structure (red) (Cα rmsd 3.7 Å). This potential is nonadditive and is inferred using the MFP on crystallographic dimer structures, ensuring that the binding processes are funneled as well (Z Luthey-Schulten, PG Wolynes *et al*., unpublished). (Figure prepared by Garegin Papoian.)

native structure means that a collective coordinate measuring distance to the native structure by itself gives a good indication of the energy of a particular configuration. The energy will decrease, on average, as we form structures that are more and more similar to the native structure of a natural protein. The energy landscape is 'funneled' [1,2,7].

The funneled organization of the energy landscape dominates the kinetics of folding. First, the funneling suggests that the process of folding will be hard to destroy by mutation as long as stability is maintained. Mild mutation will initially shift the folding routes and, if one set of routes to the native structure is completely blocked by a very destabilizing mutation, another set of routes will take over, as long as the final structure is stable. Second, funneling requires that a bottleneck must block an entire set of folding routes. Such a bottleneck explains the often-observed exponential folding kinetics and there must therefore be ensembles of structures through which the system must pass to fold. These are called 'transition state ensembles' [16]. They do not consist of saddle points on the energy surface but rather high-energy intermediate structures. The stability of transition state ensembles will largely be acquired by making use of interactions present in the native structure. Funneling therefore implies a general correlation between folding stability changes and folding rates. The funnel-like nature of the landscape also implies that the relationship between stability and rate will generally only be (logarithmically) linear over a limited range of stability change. Funneling also leads to Hammond behavior (i.e. the transition state ensemble becomes less native if the protein becomes more stable through solvent environment or temperature changes). In an extreme case, downhill folding may be expected. If folding is downhill, further increases in stability will not speed folding and, without bottlenecks, nonexponential kinetics will be expected. The simple kinetic behavior expected of a funneled landscape resembles the observed patterns of the folding of small single-domain proteins [1]. Highly frustrated systems would exhibit a variety of different kinetic traps caused by landscape ruggedness and kinetic changes under mildly different thermodynamic conditions. The theoretical study of perfectly funneled landscapes then is a good place to begin a theoretical treatment of folding kinetics. Perfect funnel or 'Go' models thus are a kind of 'perfect gas' limit for real folding. It is important to understand the dynamics implied by a perfect funnel, if only to uncover the specific effects that cause deviations [1,2].

Perfect funnel landscapes and common features of folding mechanisms

A funneled landscape is responsible for the robust ability of proteins to fold. However, a variety of detailed mechanisms may exist on a funneled energy landscape. For example, secondary structures may form before or after collapse, sidechains may order before or after the mainchain topology, one domain of a protein may fold before another. As no particular evolutionary advantage is apparent for any of these mechanisms, it is reasonable to expect to see examples of them all.

The global landscape topography is determined by evolution. Evolution works on a rather crude energy scale, making sure that folding landscapes are robust and funnel like. Disallowed mutations may cost as much as 10 k_BT. The choice of detailed folding mechanism occurs at a much finer energy scale, a little larger than the thermal fluctuations. Barriers to folding (around k_BT) are a relatively small fraction of the total binding energy (around 100 k_BT) of the native folded structure, because a good deal of the binding must be cancelled by a concomitant entropy loss (this is the renormalized energy including effects of solvation). Folding and unfolding barriers that are the result of the non-perfect cancellation between energy and configurational entropy are subtler still in energy scale, but can exceed k_BT. Thus, an experiment may seem to reveal dramatically different mechanisms for two very similar sequences that fold to the same structure (Figure 1).

Although the detailed mechanisms of folding are therefore less robust than structure, the funnel landscape has an enormous influence on the mechanism and many features of folding trajectories are common to several different detailed mechanisms. Such common patterns can be inferred from perfect funnel models — as most naturally occurring proteins have sufficiently reduced energetic frustration, the funnel landscape idea thus implies the notion, now quite well accepted as a general guideline (with exceptions), that 'topology determines folding mechanism'.

Perfect funnel or Go models include only interactions that stabilize the native structure. We discuss here coarse-grained models [17–21] and not detailed all-atom plus explicit solvent simulations [22–30]. The simplest models assume that stabilization of the native structure monotonically increases as pairs of residues are brought together (many groups). In this approximation, bottlenecks must arise from the various ways entropy and energy compensate while assembling the fold. Using a pairwise additive Cα model with a Go potential, the main structural components of the transition state ensembles of small proteins such as protein A, CI2 and SH3 were accurately reproduced [31]. Even more impressively, this model captured both the transition state ensembles and the intermediate structures of larger proteins such as barnase, RNase H, CheY, IL-1β and dihydrofolate reductase [31,32]. A landmark of such studies is the survey by Koga and Takada of 18 small proteins [33]. A similar approach has also been applied to the folding and binding

of dimers [34–36]. A study of 11 dimers predicts whether they follow the two- or three-state folding mechanism observed experimentally [35]. The folding rate of Go model proteins correlates with both the length of the protein and the relative contact order, which roughly measures the entropy cost of forming the native structure. Such a correlation was found experimentally and is even more pronounced. Of course, deliberate protein engineering can change the folding rate by a much larger amount than the entire range seen in the data set of naturally occurring proteins, so this is only a trend.

Other global trends of folding are predicted by purely analytical treatments, as reviewed by Plotkin and Onuchic [37,38]. Free energy functional schemes based on polymer theory [39–42], and elegant simplifications of these [43–45], have also proved very useful for predicting Φ-values.

Beyond the perfect funnel: devilish details of folding mechanisms

Fine-level, sequence-dependent variations occur in the kinetics of proteins with the same topology. Laboratory folding studies also reveal a greater level of specificity and granularity than predicted by the simplest landscape-based theories using pairwise additive forces. Landscape theory also allows us to understand these devilish details.

Some deviations occur because frustration should be minimized for evolutionary reasons, but there are many reasons for some residual energetic frustration to remain. Even if native contacts are stabilizing, they may be heterogeneous in magnitude. Both heterogeneity and residual energetic frustration may lead to negative Φ-values or Φ-values greater than one. Both situations have been seen in several systems. Adding inhomogeneous energies for the native contacts in homogeneous Go models predicts the existence of such sensitive frustrated sites.

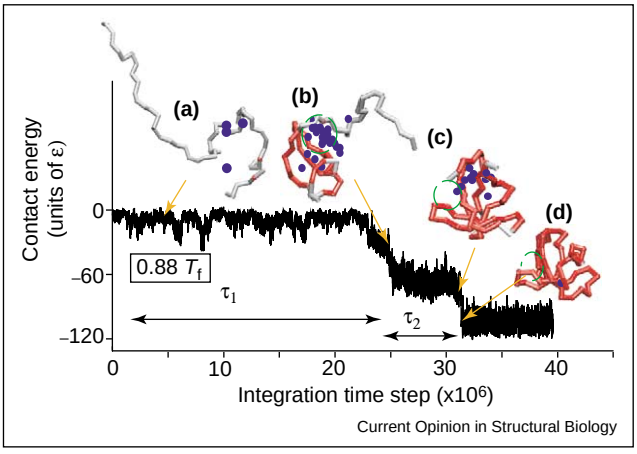
Energetic heterogeneity can bias the choice of folding routes when multiple sets of routes are allowed by symmetry. This seems to be the situation for protein L and protein G, for which the geometric near symmetry of the two structural halves of the molecules can be observed in the kinetic routes by mutations or broken by permuting the protein sequence. Small variations in sequence are able to differentiate between these mechanisms [46–50]. In particular, all-atom Go models [46], with their inherent inhomogeneity, reproduce these results.

Pairwise additive forces give lower absolute barriers to folding and faster rates than those needed to explain many laboratory measurements. If the effective inter-residue interactions involve three- or four-body terms, so that several residues need to come together for stabilization, the barriers are increased because more entropy cost must be paid before binding energy is given back. Such nonadditive forces arise from contributions of averaging over solvent or sidechain degrees of freedom. The consequences of nonadditivity have been quantified both in analytical theories [51] and in simulations of nonadditive models [52,53].

Funnel landscapes with nonadditive forces show increased levels of cooperativity and specificity for partially folded states [46,52]. This additional cooperativity makes the transition state ensemble less diffuse, more polarized and more compact. The increased compactness of the ordered regions exacerbates the effects of energetic heterogeneity. These effects are discussed in Figure 1.

Several intriguing suggestions have been made concerning the 'gating' of folding by solvent degrees of freedom. Certainly, the formation of individual contacts in a folding protein is gated by the solvent [54–56]. Garcia and

Figure 3



A typical folding trajectory for the SH3 domain represented by a minimalist Cα model with Go-like interactions generalized to include a desolvation potential. A plot of the contact energies as a function of integration time steps is shown. To have a better understanding of the kinetics, several snapshots of the chain are shown in which blue spheres are used to identify contacts separated by a single water molecule (notice that unfolded residues without blue spheres are fully solvated). We color residues with formed native contacts red to specify folded portions and residues with unformed native contacts gray to specify unfolded portions. In this trajectory, **(a)** shows an unfolded configuration in which only the short-range native contacts are formed. This early step is followed by a transition representing structural collapse to the nearly native ensemble **(b→c)**. About 23 'water' molecules are expelled during this transition. Notably, native contact formation between the diverging turn and the distal loop (indicated by a broken green circle) is crucial to this structural collapse. This region overlaps with the experimentally determined high Φ-value regions. The final transition involves the mechanism of water expulsion from the partially hydrated hydrophobic core **(c→d)**. This transition involves the formation of long-range tertiary contacts across the two sandwiched β sheets. Seventeen 'water' molecules are expelled cooperatively during this final transition. Configuration (d) has only a few residual 'water' molecules in the terminal regions (figure prepared by Margaret S Cheung).

Onuchic [57] have explained the non-Hammond pressure dependence of folding rates using this idea. The dynamical prefactors of the rate are reduced because of the necessity to expel one or more water molecules when hydrophobic contacts are made. A more extreme version of the solvent gating idea has also been mooted. Simulations of rigid objects approaching each other in solution show catastrophic drying events [58]. However, it is clear that the flexibility of partially folded chains will make such a discrete drying event less necessary, as recently demonstrated by several experiments and simulations [59–62]. Several simulations of folding show that flexible chains expel solvent molecules a few at a time (Figure 3 illustrates this fact with a minimalist model).

Conclusions

The 20th century's fixation on structure catapulted folding to center stage in molecular biology. The lessons learned about folding may, in the future, increase our understanding of many functional motions and large-scale assembly processes. The appreciation that folding physics plays a role in the allosteric function of proteins is likely to be a recurring theme in the coming years [63].

Acknowledgements

This work has been funded by the National Science Foundation (NSF)-sponsored Center for Theoretical Biological Physics (grants PHY-0216576 and 0225630), with additional support from the NSF to JNO (Grant MCB-0084797) and from the National Institutes of Health to PGW.

References

1. Onuchic JN, Luthey-Schulten Z, Wolynes PG: **Theory of protein folding: the energy landscape perspective**. *Annu Rev Phys Chem* 1997, **48**:545-600.
2. Onuchic JN, Nymeyer H, Garcia AE, Chahine J, Socci ND: **The energy landscape of protein folding: insights into folding mechanisms and scenarios**. *Adv Protein Chem* 2000, **53**:87-152.
3. Eaton WA, Munoz V, Thompson P, Chan CK, Hofrichter J: **Submillisecond kinetics of protein folding**. *Curr Opin Struct Biol* 1997, **7**:10-14.
4. Fersth AR: **Nucleation mechanisms in protein folding**. *Curr Opin Struct Biol* 1997, **7**:3-9.
5. Miranker AD, Dobson CM: **Collapse and cooperativity in protein folding**. *Curr Opin Struct Biol* 1996, **6**:31-42.
6. Bryngelson JD, Wolynes PG: **Spin-glasses and the statistical-mechanics of protein folding**. *Proc Natl Acad Sci USA* 1987, **84**:7524-7528.
7. Leopold PE, Montal M, Onuchic JN: **Protein folding funnels: kinetic pathways through a compact conformation space**. *Proc Natl Acad Sci USA* 1992, **89**:8721-8725.
8. Hardin C, Eastwood MP, Prentiss M, Luthey-Schulten Z, Wolynes PG: **Folding funnels: the key to robust protein structure prediction**. *J Comput Chem* 2002, **23**:138-146.
9. Bryngelson JD, Wolynes PG: **Intermediate and barrier crossing in a random energy model (with applications to protein folding)**. *J Phys Chem* 1989, **93**:6902-6915.
10. Udea Y, Taketomi H, Go N: **Studies of protein folding, unfolding, and fluctuations by computer simulation. 2. 3-dimensional lattice model for lysozyme**. *Biopolymers* 1978, **17**:1531-1548.
11. Socci ND, Onuchic JN: **Diffusive dynamics of the reaction coordinate for protein folding funnels**. *J Chem Phys* 1995, **103**:4732-4744.
12. Gutin A, Sali A, Abkevich V, Karplus M, Shakhnovich EI: **Temperature dependence of the folding rate in a simple protein model: search for a 'glass' transition**. *J Chem Phys* 1998, **108**:6466-6483.
13. Hardin C, Eastwood MP, Luthey-Schulten Z, Wolynes PG: **Associative memory Hamiltonians for structure prediction without homology: alpha-helical proteins**. *Proc Natl Acad Sci USA* 2000, **97**:14235-14240.
14. Hardin C, Eastwood MP, Prentiss MC, Luthey-Schulten Z, Wolynes PG: **Associative memory Hamiltonians for structure prediction without homology: alpha/beta protein**. *Proc Natl Acad Sci USA* 2003, **100**:1679-1684.
15. Jin WZ, Kambara O, Sasakawa H, Tamura A, Takada S: ***De novo* design of foldable proteins with smooth folding funnel: automated negative design and experimental verification**. *Structure* 2003, **11**:581-590.
16. Onuchic JN, Wolynes PG, Luthey-Schulten Z, Socci ND: **Towards an outline of the topography of a realistic folding funnels**. *Proc Natl Acad Sci USA* 1995, **92**:3626-3630.
17. Ozkan SB, Bahar I, Dill KA: **Transition states and the meaning of Phi-values in protein folding kinetics**. *Nat Struct Biol* 2001, **8**:765-769.
18. Levitt M, Warshel A: **Computer simulations of protein folding**. *Nature* 1975, **253**:694-698.
19. Friedrichs MS, Goldstein RA, Wolynes PG: **Generalized protein tertiary structure recognition using associative memory Hamiltonians**. *J Mol Biol* 1991, **222**:1013-1034.
20. Irback A, Peterson C, Potthast F, Sommelius O: **Local interactions and protein folding: a three-dimensional off-lattice approach**. *J Chem Phys* 1997, **107**:273-282.
21. Guo Z, Thirumalai D, Honeycutt JD: **Folding kinetics of proteins – a model study**. *J Chem Phys* 1992, **92**:525-535.
22. Boczko EM, Brooks CL III: **First principles calculation of the folding free-energy of a 3-helix bundle protein**. *Science* 1995, **269**:393-396.
23. 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.
24. Duan Y, Kollman PA: **Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution**. *Science* 1998, **282**:740-744.
25. Snow CD, Nguyen N, Pande VS, Gruebele M: **Absolute comparison of simulated and experimental protein-folding dynamics**. *Nature* 2002, **420**:102-106.
26. Garcia AE, Sanbonmatsu KY: **Alpha-helical stabilization by side chain shielding of backbone hydrogen bonds**. *Proc Natl Acad Sci USA* 2002, **99**:2782-2787.
27. Shea JE, Brooks CL III: **From folding theories to folding proteins: a review and assessment of simulation studies of protein folding and unfolding**. *Annu Rev Phys Chem* 2001, **52**:499-535.
28. Daggett V, Fersht A: **The present view of the mechanism of protein folding**. *Nat Rev Mol Cell Biol* 2003, **4**:497-502.
29. Daggett V, Fersht A: **Protein folding and unfolding at atomic resolution**. *Cell* 2002, **108**:573-582.
30. 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**:151-164.
31. Clementi C, Nymeyer H, Onuchic JN: **Topological and energetic factors: what determines the structural details of the transition state ensemble for protein folding? An investigation for small fast folding proteins**. *J Mol Biol* 2000, **298**:937-953.
32. Clementi C, Jennings PA, Onuchic JN: **How native state topology affects the folding of dihydrofolate reductase and interleukin-beta**. *Proc Natl Acad Sci USA* 2000, **97**:5871-5876.

33. Koga N, Takada S: **Roles of native topology and chain-length scaling in protein folding: a simulation study with a Go-like model**. *J Mol Biol* 2001, **313**:171-180.

34. Papoian GA, Wolynes PG: **The physics and bioinformatics of binding and folding - an energy landscape perspective**. *Biopolymers* 2003, **68**:333-349.

35. Levy Y, Onuchic JN, Wolynes PG: **Protein topology determines binding mechanism.** *Proc Natl Acad Sci USA* 2003, in press.

36. Shoemaker BA, Portman JJ, Wolynes PG: **Speeding molecular recognition by using the folding funnel: the fly-casting mechanism**. *Proc Natl Acad Sci USA* 2000, **97**:8868-8873.

37. Plotkin SS, Onuchic JN: **Understanding protein folding with energy landscape theory - Part I: basic concepts**. *Q Rev Biophys* 2002, **35**:111-167.

38. Plotkin SS, Onuchic JN: **Understanding protein folding with energy landscape theory - Part II: quantitative aspects**. *Q Rev Biophys* 2002, **35**:205-286.

39. Shoemaker BA, Wang J, Wolynes PG: **Exploring structures in protein folding funnels with free energy functionals: the transition state ensemble**. *J Mol Biol* 1999, **287**:675-694.

40. Shoemaker BA, Wang J, Wolynes PG: **Structural correlations in protein folding funnels**. *Proc Natl Acad Sci USA* 1997, **94**:777-782.

41. Portman JJ, Takada S, Wolynes PG: **Variational theory for site resolved protein folding free energy surfaces**. *Phys Rev Lett* 1998, **81**:5237-5240.

42. Portman JJ, Takada S, Wolynes PG: **Microscopic theory of protein folding rates. II. Local reaction coordinates and chain dynamics**. *J Chem Phys* 2001, **114**:5082-5096.

43. Munoz V, Eaton WA: **A simple model for calculating the kinetics of protein folding from three-dimensional structures**. *Proc Natl Acad Sci USA* 1999, **96**:11311-11316.

44. 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.

45. Galzitskaya OV, Finkelstein AV: **A theoretical search for folding/unfolding nuclei in three-dimensional protein structures**. *Proc Natl Acad Sci USA* 1999, **96**:11299-11304.

46. Clementi C, Garcia AE, Onuchic JN: **All-atom representation study of protein L**. *J Mol Biol* 2003, **326**:933-954.

47. 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.

48. Amato NM, Dill KA, Song G: **Using motion planning to map protein folding landscapes and analyze folding kinetics of known native structures**. *J Comput Biol* 2003, **10**:239-255.

49. Karanicolas J, Brooks CL III: **The origins of asymmetry in the folding transition states of protein L and protein G**. *Protein Sci* 2002, **11**:2351-2361.

50. Sorenson JM, Head-Gordon T: **Matching simulation and experiment: a new simplified model for simulating protein folding**. *J Comput Biol* 2000, **7**:469-481.

51. Plotkin SS, Wang J, Wolynes PG: **Correlated energy landscape model for finite, random heteropolymers**. *Phys Rev E* 1996, **53**:6271-6296.

52. Kaya H, Chan HS: **Solvation effects and driving forces for protein thermodynamic and kinetic cooperativity: how adequate is native-centric topological modeling?** *J Mol Biol* 2003, **326**:911-931.

53. Hao MH, Scheraga HA: **Molecular mechanisms for cooperative folding of proteins**. *J Mol Biol* 1998, **277**:973-983.

54. Honeycutt JD, Thirumalai D: **The nature of folded states of globular proteins**. *Biopolymers* 1992, **32**:695-709.

55. Hummer G, Garde S, Garcia AE, Paulaitis ME, Pratt LR: **The pressure dependence of hydrophobic interactions is consistent with the observed pressure denaturation of proteins**. *Proc Natl Acad Sci USA* 1998, **95**:1552-1555.

56. Sorenson JM, Head-Gordon T: **Toward minimalist models of larger proteins: a ubiquitin-like protein**. *Proteins* 2002, **46**:368-379.

57. Hillson N, Onuchic JN, Garcia AE: **Pressure-induced protein-folding/unfolding kinetics**. *Proc Natl Acad Sci USA* 1999, **96**:14848-14853.

58. ten Wolde PR, Chandler D: **Drying-induced hydrophobic polymer collapse**. *Proc Natl Acad Sci USA* 2002, **99**:6539-6543.

59. Cheung MS, Garcia AE, Onuchic JN: **Protein folding mediated by solvation: water expulsion and formation of the hydrophobic core occurs after structural collapse**. *Proc Natl Acad Sci USA* 2002, **99**:685-690.

60. Shea JE, Onuchic JN, Brooks CL III: **Probing the free energy landscape of the src-SH3 protein domain**. *Proc Natl Acad Sci USA* 2002, **99**:16064-16068.

61. Fernandez-Escamilla AM, Cheung MS, Vega MC, Wilmanns M, Onuchic JN, Serrano L: **Solvation in protein folding analysis, combination of theoretical and experimental approaches.** *Proc Natl Acad Sci USA* 2003, in press.

62. Garcia AE, Onuchic JN: **Folding a protein in a computer an atomic description of the folding/unfolding of protein A**. *Proc Natl Acad Sci USA* 2003, **100**:13898-13903.

63. Miyashita O, Onuchic JN, Wolynes PG: **Nonlinear elasticity, proteinquakes, and the energy landscapes of functional transitions in proteins**. *Proc Natl Acad Sci USA* 2003, **100**:12570-12575.