Full text provided by www.sciencedirect.com
SCIENCE DIRECT

Network and graph analyses of folding free energy surfaces

Amedeo Caflisch

Protein folding is governed by a complex free energy surface whose entropic contributions are relevant because of the large number of degrees of freedom involved. Such complexity, in particular the conformational heterogeneity of the denatured state, is hidden in projections onto one or two order parameters (e.g. fraction of native contacts and/or radius of gyration), which usually results in relatively smooth surfaces. Recent approaches borrowed from network and graph theory have yielded quantitative unprojected representations of the free energy surfaces of a β-hairpin and a three-stranded β-sheet peptide using equilibrium folding-unfolding molecular dynamics simulations. Interestingly, the network and graph analyses of these structured peptides have revealed a very heterogeneous denatured state ensemble. It includes high-enthalpy, high-entropy conformations with fluctuating non-native secondary structure, as well as low-enthalpy, low-entropy traps.

Addresses
Department of Biochemistry, University of Zurich, CH-8057 Zurich, Switzerland

Corresponding author: Caflisch, Amedeo (caflisch@bioc.unizh.ch)

Current Opinion in Structural Biology 2006, **16**:71–78

This review comes from a themed issue on
Folding and binding
Edited by Mikael Oliveberg and Eugene I Shakhnovich

Available online 18th January 2006

0959-440X/$ – see front matter
© 2005 Elsevier Ltd. All rights reserved.

DOI 10.1016/j.sbi.2006.01.002

Introduction

From the almost concomitant formation of native-like interactions observed using different experimental techniques, the folding of many small single-domain proteins appears to be an effectively two-state reaction [1], that is to say, only the native and denatured states are populated at equilibrium. Yet, protein folding is a complex process driven by non-covalent (van der Waals and electrostatic) interactions involving the atoms of the protein, the surrounding water molecules and ions. During folding, the loss of entropy in the protein chain is counterbalanced by favorable interactions between protein atoms; the former is thought to be mainly responsible for the activation barrier. The major role played by the entropic contributions indicates that analysis of the free energy surface of proteins is more important than analysis of the potential energy surface, in particular for investigating the folding process [2–5]. The potential energy surfaces of several molecular systems, ranging from peptides [6] to a virus capsid [7], have been characterized, but analysis of the free energy surface is much more difficult.

One common way to investigate and display the folding free energy landscape is to study it as a function of one or more order parameters, that is, suitably chosen macroscopic quantities that should discriminate the different states of the protein. In this context, it is useful to distinguish between an 'order parameter' and a 'reaction coordinate'. The former is any observable that identifies configurations in different stable states. A reaction coordinate describes the complete process or reaction, that is, it must correlate closely with the location of every state and identify the transition state(s) along the folding pathway(s). Hence, a reaction coordinate is a good order parameter, but the inverse is not necessarily true. For example, it is common in the study of protein folding to use the fraction of native contacts (Q) [8]. Q is a satisfactory reaction coordinate for Go-model proteins [9], in which favorable interactions occur only between residues that are in contact in the folded state [10]. On the other hand, for transferable potentials (e.g. those based on physico-chemical principles, such as AMBER, CHARMM and OPLS), Q describes well only the fully folded ($Q = 1$) and unfolded ($Q = 0$) states. In fact, for a structured peptide simulated by a transferable force-field, some conformations with $Q \sim 0.7$ were found to belong to the denatured state ensemble and conformations with $Q \sim 0.3$ to the folded state [11]. Similar conclusions on the inadequacy of Q were drawn on the basis of lattice model simulation results [12].

Free energy projections onto (often naively chosen) order parameters have been used to analyze many aspects of protein folding. Stable states are usually associated with local free energy minima on the projected landscape. The depth of the minima is considered proportional to the stability of the states and the barriers between different minima indicate activation energies between states. In many cases, this approach reveals a surprisingly simple two-state picture of protein folding, which, although deceptively similar to the interpretation of the experimental observations, is in striking contrast to the complexity of the actual free energy surface. Hence, using free energy projections for the study of the kinetics of protein folding requires knowledge of an appropriate reaction coordinate. Such a reaction coordinate is not easily accessible and/or identifiable [12]. Given the complexity of protein folding, which, as mentioned at the beginning, originates from the large number of degrees of

freedom and pairs of interacting atoms, a simplified description is needed (e.g. to compare with experimental data), even though it might miss essential aspects of the process [8].

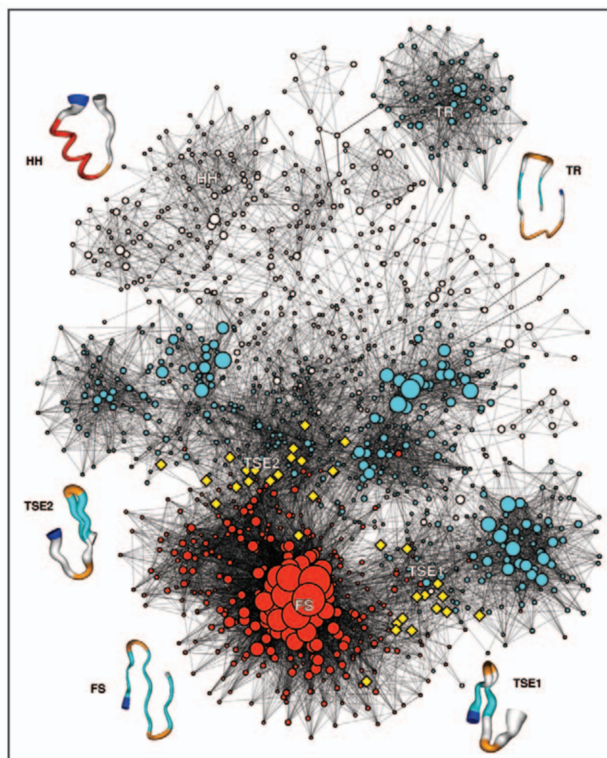
The focus of this review is on recent approaches to investigate the free energy surface governing the folding of structured peptides and small proteins. The emphasis is on methods that do not use projections onto arbitrarily chosen order parameters. Such methods have provided interesting insights into the elusive folding transition state and the denatured state ensemble. Most of these approaches and their applications were published in 2004–2005, and are still the subject of intense investigations.

Conformational space networks

Many complex systems, such as social interactions, the Internet, metabolic pathways [13,14] and protein structures [15–17], have been modeled as networks. Intriguingly, common topological properties have emerged from their organization [18]. Scala *et al.* [19] have mapped the conformational space of a short two-dimensional lattice polymer chain onto a network in which a link between two nodes indicates their interconversion in a single Monte Carlo move of the chain. Doye [20,21] has investigated the potential energy landscape of a Lennard–Jones cluster of atoms by dividing the surface into basins of attraction surrounding minima and linking those basins that are directly connected by a saddle point.

We have used complex network analysis [22] to study the conformational space and folding of a designed 20-residue peptide (β3s), whose solution conformation had been previously investigated by NMR spectroscopy [23] and implicit solvent molecular dynamics simulations [24,25]. In the latter, β3s was shown to reversibly fold, irrespective of the starting conformation, to the NMR solution conformation, a three-stranded antiparallel β-sheet. The short sequence of the peptide and the implicit treatment of the solvent allow the sampling of several folding-unfolding events at the melting temperature of 330 K along a 1 μs trajectory in about 10 days on a single processor of a personal computer. To build the network, the β3s conformations and transitions sampled by molecular dynamics were considered as nodes and links, respectively [22]. In this way, free energy minima and their connectivity emerge without requiring projections onto arbitrarily chosen reaction coordinates (Figure 1). As previously observed for a variety of networks as diverse as the Internet and the protein interactions within a cell, the conformational space network of polypeptide chains is a scale-free network, that is, the distribution of the number of possible connections of a conformation follows a power law. The reasons for the power law distribution are not clear. Interestingly, a correlation was found between the statistical weight (size of the node) and connectivity (number of links to a node) — the most connected nodes are also low-lying minima on the free energy landscape. Another observation was that the native basin of the structured peptide shows a hierarchical organization of conformations. Such an organization was not observed for a random heteropolymer that lacks a native state (i.e. a predominant free energy minimum) [22].

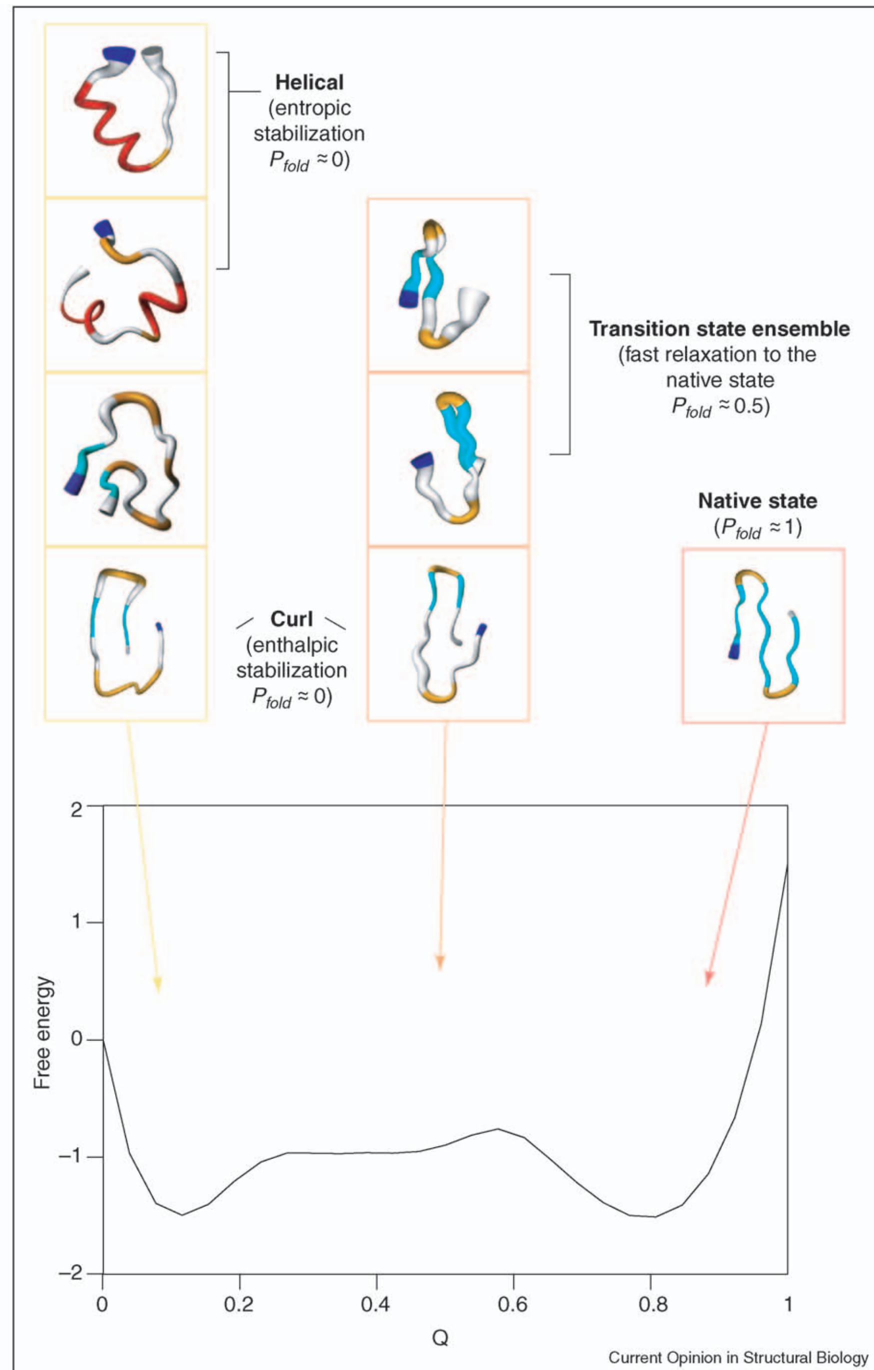
Figure 1



Conformational space network of the designed three-stranded antiparallel β-sheet peptide β3s. Nodes represent conformations and links represent transitions between them, as sampled during 10 μs implicit solvent molecular dynamics simulations at the melting temperature of 330 K. The size and color of the nodes reflect the statistical weight and average neighbor connectivity, respectively [22]. Representative conformations are shown by a pipe colored according to secondary structure: white for coil, red for α-helix, orange for turn or bend, cyan for β-strand and blue for the N-terminus. The variable radius of the pipe reflects the structural variability of the snapshots within a node. The yellow diamonds are folding transition state conformations. HH, TR, TSE and FS are the helical, trap, transition state ensemble and folded states, respectively. Reproduced with permission from [22].

One of the most interesting findings of the network analysis was that the denatured state ensemble of β3s is very heterogeneous [22] (see below). The complexity of the denatured state is hidden in the projection of the free energy onto Q (Figure 2) or onto two order parameters describing the N- and C-terminal hairpins (see Figure 5 of [25]). Both projections yield a smooth profile. Using the root mean square deviation (rmsd) from the folded structure or the radius of gyration as reaction coordinates is even less selective [26]. The usefulness

Figure 2



Projection of the free energy of β3s onto the fraction of native contacts (Q). The profile of the projected free energy is smooth, in striking contrast to the network representation of Fig. 1; both figures are based on the same 10 μs sampling. For values of $Q < 0.8$, the projection masks the complexity of the non-native states, that is to say, it groups together structurally different conformations, for example, in the transition state (orange boxes) and denatured state ensemble (yellow boxes). Moreover, kinetically different conformations, such as the two transition state representatives (p_{fold} ~0.5, top two orange boxes) and the curl-like trap (p_{fold} ~0, bottom orange box), are not separated by Q (i.e. both have Q ~0.5).

of the network analysis was also shown by employing some of the network properties to identify transition state conformations and two main average folding pathways [22]. To date, network analysis has mainly focused on the study of network topology (i.e. the connectivity between different conformations and the overall shape of the free energy surface) and has not yet been used to determine transition rates and kinetics.

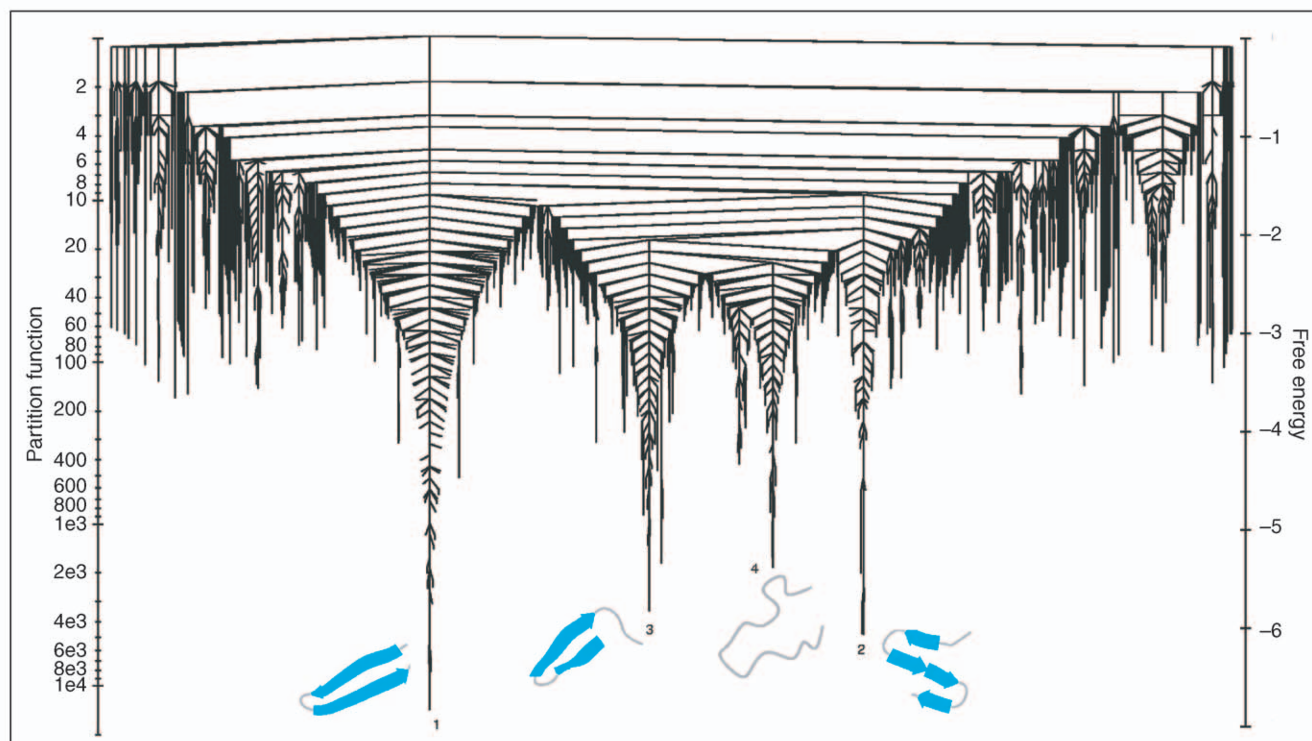
Recently, Andrec *et al.* [27] have combined replica exchange molecular dynamics simulations and a network model to investigate the kinetics of peptide folding. Pairs of snapshots saved along 20 parallel trajectories (at temperatures between 270 K and 690 K) were connected if they had similar Cα–Cα distance matrices and were sampled in runs at the same or nearest temperature values. With this combined approach, they observed that the coil-to-hairpin transition of a 16-residue peptide (the C-terminal segment from the B1 domain of protein G) often proceeds through metastable helical conformations [27]. An unanswered question concerns the use of the replica exchange technique for kinetic analysis, which is not justified because the peptide replicas are periodically swapped through different thermodynamic states (i.e. from a high-temperature replica to a low-temperature replica).

Transition disconnectivity graphs

Because of the high complexity of the potential energy surface of molecular systems, Becker and Karplus [6] introduced a pictorial, but quantitative, description that is based on the disconnectivity graph (DG). A DG provides a visualization based on a dendrogram of all of the minimum-saddle-minimum triads. Such tree-like graphs have been already applied to a variety of molecular systems [7].

Because the equilibrium and dynamic properties of peptides and proteins depend on the free energy, rather than the potential energy, the DG approach has been recently extended by Krivov and Karplus [28] to analyze the free energy surface. They have developed the transition DG (TRDG) approach, which exploits an isomorphism between the total rate connecting two free energy minima (considering all possible pathways) and the maximum flow through a network with capacitated edges (i.e. edges directly or indirectly connecting two nodes and having a certain flow capacity). The TRDG was determined first for the alanine hexapeptide and the Arg-Gly-Asp-Ser peptide [28], and more recently for the 16-residue β-hairpin peptide from the B1 domain of protein G [29]. A very interesting finding of the TRDG analysis of the β-hairpin peptide is that the denatured basin is not funnel like (Figure 3) and, in spite of the traps, which equilibrate rapidly (see below), exponential folding behavior is observed. In this context, it is important to underline that the original funnel diagram [30] plots along the vertical axis the effective energy, which is the sum of the intrapeptide energy and solvation free energy [31]. However, it is the

Figure 3



Transition disconnectivity graph of a β-hairpin (the C-terminal segment from the B1 domain of protein G). A total of 4 μs implicit solvent molecular dynamics simulations at 360 K were sampled to obtain a sufficient number of folding-unfolding events [29]. Representative structures of the deepest free energy minima are shown and labeled 1–4. The left vertical axis shows the partition function of the minima and barriers. The right vertical axis shows the free energy of the minima and barriers. Reproduced with permission from [29].

total free energy — the effective energy plus conformational entropy (the latter being the horizontal width of the funnel) — that determines the folding process [2,3]. As found for β3s, projections of the β-hairpin free energy surface onto two dimensions (e.g. rmsd from the folded structure and radius of gyration, or the two most important principal components; Figure 4 of [29]) are smooth and hide the complexity of the denatured state.

To obtain a simplified description of the folding process, the complex form of the TRDG of the β-hairpin was 'reduced' to a network of only five nodes (i.e. free energy basins) by preserving the barriers between basins [29]. It is clear from the simplified network that most folding transitions go through the 'entropic' basin [29]. (The term 'entropic' originates from the observation that the free energy of this basin has a significant contribution from the peptide conformational entropy. In other words, the entropic basin consists of many shallow sub-basins that could not be separated because all partitions are closely connected by small equilibration times, i.e. 1 ns or less.) As a consequence, the three non-native basins that do not belong to the entropic basin appear to be off-pathway. Although the choice of the number of nodes for building the simplified network was not difficult in the case of the β-hairpin, which has four pronounced basins and the entropic basin (Figure 3), it might not always be simple and might involve the choice of an arbitrary threshold for more complicated TRDGs.

The essential element of both the conformational space network [22] and TRDG [29] approaches is the description of free energy minima and basins, not according to geometrical characteristics (such as Q or rmsd from the folded structure), but rather according to the transitions at equilibrium. In other words, given an equilibrium folding-unfolding trajectory, the population of the states provides the relative free energies. Moreover, in the case of the TRDG, the rate of transitions between states yields the free energy barriers, which are illustrated by the TRDG dendrogram. It is important to underline that both the network and TRDG approaches require a clusterization to group snapshots saved along equilibrium trajectories into states (i.e. the nodes of the network [22] or vertices of the TRDG dendrogram [29]). Some evidence of the robustness upon changing clusterization algorithm (based on rmsd or secondary structure string) has been provided [22,29], but this issue has to be analyzed in more detail. Several investigations of the effects of clustering (i.e. grouping snapshots into macrostates) have been published recently [32,33].

Heterogeneity of the denatured state ensemble

The most surprising result obtained from the network analysis of β3s and the graph analysis of the protein G β-hairpin is the heterogeneity of the denatured state ensemble. In fact, it includes a variety of high-enthalpy, high-entropy conformations, for example, the partially helical conformations of β3s (denoted HH in Figure 1), as well as low-enthalpy, low-entropy structures, for example, the curl-like trap (TR) [22]. The former are loosely linked clusters of conformations with non-native secondary structure (see Table 1 in [22]), and are characterized by an unfavorable effective energy (the sum of peptide potential energy and solvation energy) and fluctuating unstructured residues (e.g. the terminus of the helix shown at the top left of Figure 1). In contrast, low-enthalpy, low-entropy traps form tightly linked clusters with almost identical secondary and tertiary structure, and have a favorable effective energy (similar to or even slightly more favorable than that of the native structure) and no fluctuating residues (e.g. Figure 1, top right). Taken together, these results indicate that the folded state of these two β-structured peptides is entropically favored over low-enthalpy non-native conformations, such as the curl-like trap, that is to say, the folded state has more flexibility than the curl-like trap. It is not possible to generalize the entropic stability of the native state, which could originate from approximations inherent to the implicit solvent model, in particular the treatment of charged groups [24]. A possible explanation for the observation that the model favors a misfolded conformation of β3s is that the C-terminal carboxy is involved in four hydrogen bonds in the trap (with the backbone NHs of residues 4–7), whereas both termini undergo relatively large fluctuations in the folded state [22].

The TRDG analysis of the β-hairpin also revealed the hidden complexity of its denatured state ensemble. In fact, besides the aforementioned entropic basin, several deep sub-basins with low enthalpy and low entropy were observed in the TRDG, although they are not evident on projected surfaces [29]. Some of these sub-basins (e.g. β-hairpin with shifted turn, see conformation 3 in Figure 3) were shown to have a more favorable effective energy than the folded state, which is stabilized by its higher conformational entropy. More recently, Chekmarev *et al.* [34] have used the TRDG method to investigate a 27-residue lattice heteropolymer subject to Monte Carlo dynamics on a cubic lattice [34]. The authors report that the denatured basin of the 27-residue heteropolymer consists of several deep sub-basins (with low enthalpy and low entropy) that are separated by high barriers; no entropic basin was mentioned.

The heterogeneity of the denatured state ensembles of β3s and the β-hairpin, in particular the statistically significant weight of non-native (i.e. helical) conformations of β3s, is in contrast to the observation of native-like mean structure in the denatured state ensembles of a small helical protein and two structured peptides, as found by distributed computing [35]. It is likely that the denatured state ensemble is not correctly sampled

in the distributed computing approach, in which simulations are usually started from a fully extended conformation that rapidly collapses without allowing proper equilibration within the unfolded state. Moreover, the fastest pathways are not necessarily the most probable ones, as has been suggested by Fersht [36] and confirmed by molecular dynamics simulations [37,38].

Reaction coordinates for the identification of the transition state

Given the lack of discriminatory ability of simple order parameters (e.g. Q, the fraction of native contacts, as mentioned above), more sophisticated approaches have been developed to investigate the folding transition state. The folding probability (p_{fold}) of a protein snapshot saved along a Monte Carlo or molecular dynamics trajectory is the probability to fold before unfolding [12]. It is a useful measure of kinetic distance from the folded state and can be used to validate transition state ensemble structures, which should have $p_{\text{fold}} = 0.5$. Such validation consists of starting a large number of trajectories from putative transition state structures with different initial atomic velocities and counting the number of those that fold within a 'commitment' time; this time has to be much longer than the shortest timescales of conformational fluctuations and much shorter than the average folding time. The concept of the p_{fold} calculation originates from a method for determining transmission coefficients, starting from a known transition state [39] and the identification of simpler transition states in protein dynamics (e.g. tyrosine ring flips) [40]. The approach has been used to identify the otherwise very elusive folding transition state ensemble by means of atomistic Monte Carlo off-lattice simulations of small proteins with a native-centric Go potential [41–43], as well as implicit solvent molecular dynamics [22,44] and Monte Carlo [45] simulations with a physico-chemical-based potential. The number of trial simulations needed for the reliable evaluation of p_{fold} makes the estimation of the folding probability computationally very expensive. For this reason, we have introduced a method to estimate the folding probabilities of all structures visited along an equilibrium folding-unfolding trajectory without any additional simulation [46]. The method requires structural clustering and, for each cluster, counting the fraction of snapshots that proceed to the folded state before unfolding. This method has been applied to the β3s peptide and to a set of 32 single-point mutants thereof to successfully extract Φ-values [47] from equilibrium folding-unfolding simulations [11]. It was found that Φ-values calculated from folding and unfolding rates measured along the trajectories are reliable if the stability loss upon mutation is larger than about 0.6 kcal/mol, in agreement with a suggestion based on experimental data [48]. Moreover, the molecular dynamics and p_{fold} analyses of β3s and its mutants revealed the presence of specific non-native interactions at the transition state for most of the peptides [11].

Recently, Best and Hummer [9] have proposed a variational procedure to optimize reaction coordinates using snapshots from transition path sampling or equilibrium folding-unfolding trajectories. For a Go-like Cα model of a three-helix bundle protein, a Monte Carlo procedure was used to optimize the weights of a projection onto the matrix of all (native and non-native) contacts, such that transition state structures were condensed into a single peak of the probability of being on a transition path. Starting from an initial matrix with uniform weights, they obtained a projection that accurately located the transition state ensemble, as verified by traditional p_{fold} analysis. In their model, helices 2 and 3 are fully folded at the transition state, whereas helix 1 is only partially structured and is not properly docked against the scaffold consisting of helices 2 and 3. Despite the successful application to the Go-like Cα model, Best and Hummer [9] conclude their article with the cautionary note that an appropriate basis set for the variational optimization of reaction coordinates might be difficult to guess and might require the inclusion of solvent coordinates.

To identify reaction coordinates for complex systems that include water degrees of freedom, Ma and Dinner [49] have employed an automatic procedure that requires a set of structures distributed uniformly with respect to p_{fold}. Given such a database of structures and their p_{fold} values, neural networks are used to determine the functional dependence of p_{fold} on sets of coordinates, such as backbone dihedral angles and interatomic distances. A genetic algorithm selects the combination of coordinates that are given as input to the neural network to yield the best fit. The hybrid genetic algorithm-neural network approach had been developed for quantitative structure/activity relationship studies [50] and also employed to investigate the folding ability of lattice models of proteins [51]. The approach of Ma and Dinner uses p_{fold}, a reaction coordinate that provides a quantitative description of the dynamic behavior of every state along a trajectory, to guide a rather complex optimization procedure through the space of physically meaningful variables. These variables yield mechanistic insights into the process under study because they relate directly to quantities that can be measured and varied computationally, whereas p_{fold} *per se* is not informative. Application to the $C_{7eq} \rightarrow \alpha_R$ conformational isomerization of the alanine dipeptide in the presence of explicit water molecules successfully identified a set of variables that specify the transition state [49]. It will be interesting to verify if the optimization procedure of Ma and Dinner will find useful applications in the study of small protein folding, for which the number of degrees of freedom of the solute is much larger than that of the alanine dipeptide.

Conclusions

Recently, approaches borrowed from topological [22] and flow [29] analyses of networks (whereby the nodes and

links represent conformations and transitions between them, respectively) have been adapted to obtain unprojected graphical representations of and quantitative insights into the free energy surfaces of structured peptides. The striking contrast between the complexity of the free energy surface and its projection onto one or two order parameters provides a cautionary note for the interpretation of the latter. In particular, the heterogeneity of the denatured state ensemble of structured peptides (with fluctuating, i.e. entropically stabilized, non-native secondary structure), which has emerged from the network [22] (Figure 1) and TRDG [29] (Figure 3) analyses, is hidden in projections of the free energy surface onto order parameters, such as Q (Figure 2), or geometrical progress variables, such as rmsd and/or radius of gyration. The relevance of such heterogeneity is related to its possible experimental manifestations, which is a challenge for experimentalists.

Efficient methods have also been developed recently to determine the probability of folding (p_{fold}) [46] or optimize a reaction coordinate [9,49] to identify conformations in the transition state ensemble. All of these methods require the availability of several folding-unfolding transitions, which can be sampled by implicit solvent models for structured peptides and Go models for small proteins. The former have allowed the investigation of the importance of non-native interactions in the transition state [11] and denatured state ensemble [22,29], a type of analysis for which native-centric Go models are not appropriate.

Finally, it remains to be determined whether the entropic stabilization of the folded state, with respect to non-native traps with very favorable intraprotein interactions, is a peculiarity of some structured peptides (e.g. the β-sheet and β-hairpin discussed in this review) or could play a role also in (small) proteins, for example, those with flexible terminal segments and/or loops (such as chymotrypsin inhibitor 2).

Acknowledgements

I am grateful to Martin Karplus, Sergei Krivov, Stefanie Muff, Emanuele Paci, Francesco Rao and Gianni Settanni for very useful discussions, and Eugene Shakhnovich for interesting comments on this manuscript. The simulations of β3s were performed on Matterhorn, a Beowulf Linux cluster at the Informatikdienste of the University of Zurich, and I thank C Bolliger, T Steenbock and A Godknecht for installing and maintaining the Linux cluster. This work was supported by a grant from the Swiss National Science Foundation.

References

1. Jackson SE: **How do small single-domain proteins fold?** *Fold Des* 1998, **3**:R81-R91.
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.
3. Mirny L, Shakhnovich E: **Protein folding theory: from lattice to all-atom models**. *Annu Rev Biophys Biomol Struct* 2001, **30**:361-396.
4. Daggett V, Fersht A: **The present view of the mechanism of protein folding**. *Nat Rev Mol Cell Biol* 2003, **4**:497-502.
5. Wolynes PG: **Energy landscapes and solved protein-folding problems**. *Philos Transact A Math Phys Eng Sci* 2005, **363**:453-464.
6. Becker OM, Karplus M: **The topology of multidimensional potential energy surfaces: theory and application to peptide structure and kinetics**. *J Chem Phys* 1997, **106**:1495-1517.
7. Wales DJ: **The energy landscape as a unifying theme in molecular science**. *Philos Transact A Math Phys Eng Sci* 2005, **363**:357-375.
8. Chan HS, Dill KA: **Protein folding in the landscape perspective: chevron plots and non-Arrhenius kinetics**. *Proteins* 1998, **30**:2-33.
9. Best RB, Hummer G: **Reaction coordinates and rates from transition paths**. *Proc Natl Acad Sci USA* 2005, **102**:6732-6737.
10. Takada S: **Go-ing for the prediction of protein folding mechanisms**. *Proc Natl Acad Sci USA* 1999, **96**:11698-11700.
11. Settanni G, Rao F, Caflisch A: **Φ-Value analysis by molecular dynamics simulations of reversible folding**. *Proc Natl Acad Sci USA* 2005, **102**:628-633.
12. Du R, Pande VS, Grosberg AY, Tanaka T, Shakhnovich ES: **On the transition coordinate for protein folding**. *J Chem Phys* 1998, **108**:334-350.
13. Barabasi AL, Oltvai ZN: **Network biology: understanding the cell's functional organization**. *Nat Rev Genet* 2004, **5**:101-113.
14. Palla G, Derenyi I, Farkas I, Vicsek T: **Uncovering the overlapping community structure of complex networks in nature and society**. *Nature* 2005, **435**:814-818.
15. Dokholyan NV, Shakhnovich B, Shakhnovich EI: **Expanding protein universe and its origin from the biological Big Bang**. *Proc Natl Acad Sci USA* 2002, **99**:14132-14136.
16. Vendruscolo M, Dokholyan NV, Paci E, Karplus M: **Small-world view of the amino acids that play a key role in protein folding**. *Phys Rev E* 2002, **65**:061910.
17. Greene LH, Higman VA: **Uncovering network systems within protein structures**. *J Mol Biol* 2003, **334**:781-791.
18. Newman MEJ: **The structure and function of complex networks**. *Siam Rev* 2003, **45**:167-256.
19. Scala A, Amaral LAN, Barthelemy M: **Small-world networks and the conformation space of a short lattice polymer chain**. *Europhys Lett* 2001, **55**:594-600.
20. Doye JP: **Network topology of a potential energy landscape: a static scale-free network**. *Phys Rev Lett* 2002, **88**:238701.
21. Doye JP, Massen CP: **Characterizing the network topology of the energy landscapes of atomic clusters**. *J Chem Phys* 2005, **122**:84105.
22. Rao F, Caflisch A: **The protein folding network**. *J Mol Biol* 2004, **342**:299-306.
23. De Alba E, Santoro J, Rico M, Jimenez MA: ***De novo* design of a monomeric three-stranded antiparallel beta-sheet**. *Protein Sci* 1999, **8**:854-865.
24. Ferrara P, Apostolakis J, Caflisch A: **Evaluation of a fast implicit solvent model for molecular dynamics simulations**. *Proteins* 2002, **46**:24-33.
25. Ferrara P, Caflisch A: **Folding simulations of a three-stranded antiparallel beta-sheet peptide**. *Proc Natl Acad Sci USA* 2000, **97**:10780-10785.
26. Ferrara P, Caflisch A: **Native topology or specific interactions: what is more important for protein folding?** *J Mol Biol* 2001, **306**:837-850.
27. Andrec M, Felts AK, Gallicchio E, Levy RM: **Protein folding pathways from replica exchange simulations and a kinetic network model**. *Proc Natl Acad Sci USA* 2005, **102**:6801-6806.

28. Krivov SV, Karplus M: **Free energy disconnectivity graphs: application to peptide models**. *J Chem Phys* 2002, **117**:10894-10903.

29. Krivov SV, Karplus M: **Hidden complexity of free energy surfaces for peptide (protein) folding**. *Proc Natl Acad Sci USA* 2004, **101**:14766-14770.

30. 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.

31. Cavalli A, Haberthur U, Paci E, Caflisch A: **Fast protein folding on downhill energy landscape**. *Protein Sci* 2003, **12**:1801-1803.

32. Swope WC, Pitera JW, Suits F: **Describing protein folding kinetics by molecular dynamics simulations. 1. Theory**. *J Phys Chem B* 2004, **108**:6571-6581.

33. Singhal N, Snow CD, Pande VS: **Using path sampling to build better Markovian state models: predicting the folding rate and mechanism of a tryptophan zipper beta hairpin**. *J Chem Phys* 2004, **121**:415-425.

34. Chekmarev SF, Krivov SV, Karplus M: **Folding time distributions as an approach to protein folding kinetics**. *J Phys Chem B* 2005, **109**:5312-5330.

35. Zagrovic B, Snow CD, Khaliq S, Shirts MR, Pande VS: **Native-like mean structure in the unfolded ensemble of small proteins**. *J Mol Biol* 2002, **323**:153-164.

36. 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.

37. Paci E, Cavalli A, Vendruscolo M, Caflisch A: **Analysis of the distributed computing approach applied to the folding of a small beta peptide**. *Proc Natl Acad Sci USA* 2003, **100**:8217-8222.

38. Marianayagam NJ, Fawzi NL, Head-Gordon T: **Protein folding by distributed computing and the denatured state ensemble**. *Proc Natl Acad Sci USA* 2005, **102**:16684-16689.

39. Chandler D: **Statistical-mechanics of isomerization dynamics in liquids and transition-state approximation**. *J Chem Phys* 1978, **68**:2959-2970.

40. Northrup SH, Pear MR, Lee CY, McCammon JA, Karplus M: **Dynamical theory of activated processes in globular-proteins**. *Proc Natl Acad Sci USA* 1982, **79**:4035-4039.

41. Li L, Shakhnovich EI: **Constructing, verifying, and dissecting the folding transition state of chymotrypsin inhibitor 2 with all-atom simulations**. *Proc Natl Acad Sci USA* 2001, **98**:13014-13018.

42. Hubner IA, Shimada J, Shakhnovich EI: **Commitment and nucleation in the protein G transition state**. *J Mol Biol* 2004, **336**:745-761.

43. Hubner IA, Edmonds KA, Shakhnovich EI: **Nucleation and the transition state of the SH3 domain**. *J Mol Biol* 2005, **349**:424-434.

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

45. Lenz P, Zagrovic B, Shapiro J, Pande VS: **Folding probabilities: a novel approach to folding transitions and the two-dimensional Ising-model**. *J Chem Phys* 2004, **120**:6769-6778.

46. Rao F, Settanni G, Guarnera E, Caflisch A: **Estimation of protein folding probability from equilibrium simulations**. *J Chem Phys* 2005, **122**:184901.

47. Fersht AR, Matouschek A, Serrano L: **The folding of an enzyme.1. Theory of protein engineering analysis of stability and pathway of protein folding**. *J Mol Biol* 1992, **224**:771-782.

48. Fersht AR, Sato S: **Φ-Value analysis and the nature of protein-folding transition states**. *Proc Natl Acad Sci USA* 2004, **101**:7976-7981.

49. Ma A, Dinner AR: **Automatic method for identifying reaction coordinates in complex systems**. *J Phys Chem B* 2005, **109**:6769-6779.

50. So SS, Karplus M: **Evolutionary optimization in quantitative structure-activity relationship: an application of genetic neural networks**. *J Med Chem* 1996, **39**:1521-1530.

51. Dinner AR, So SS, Karplus M: **Use of quantitative structure-property relationships to predict the folding ability of model proteins**. *Proteins* 1998, **33**:177-203.