

# Is there a unifying mechanism for protein folding?

Valerie Daggett<sup>1</sup> and Alan R. Fersht<sup>2</sup>

<sup>1</sup>Department of Medicinal Chemistry, University of Washington, Seattle, WA 98195-7610, USA

<sup>2</sup>Department of Chemistry and MRC Center for Protein Engineering, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK

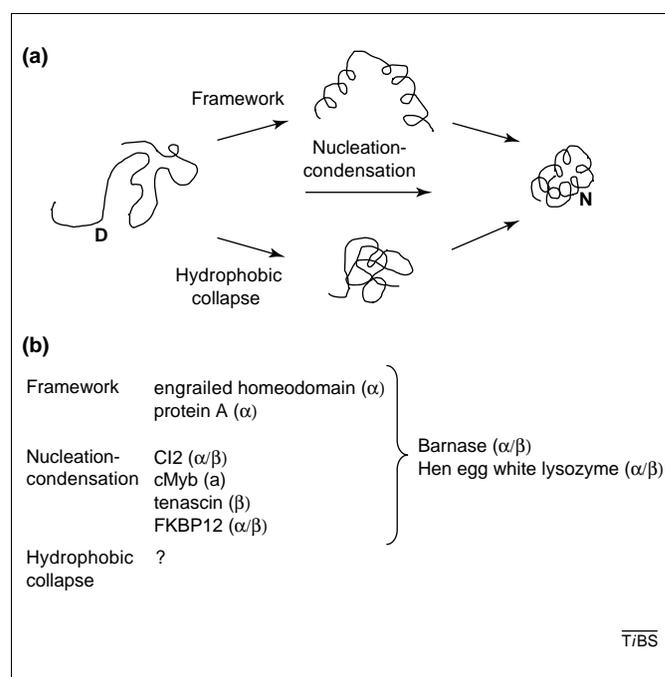
**Proteins appear to fold by diverse pathways, but variations of a simple mechanism – nucleation-condensation – describe the overall features of folding of most domains. In general, secondary structure is inherently unstable and its stability is enhanced by tertiary interactions. Consequently, an extensive interplay of secondary and tertiary interactions determines the transition-state for folding, which is structurally similar to the native state, being formed in a general collapse (condensation) around a diffuse nucleus. As the propensity for stable secondary structure increases, folding becomes more hierarchical and eventually follows a framework mechanism where the transition state is assembled from pre-formed secondary structural elements.**

Protein folding through random, unbiased searching of all possible conformations cannot occur; such a mechanism would take an infinite amount of time [1]. Thus, there appear to be pathways that simplify the mechanism of folding by breaking it down into sequential steps. The search for the holy grail of a single, simple mechanism of protein folding led to a multitude of proposals. The first – nucleation-growth [2] – proposed that tertiary structure propagates rapidly from an initial nucleus of local secondary structure. However, nucleation dropped from favor as it predicts the absence of folding intermediates, and the field of protein folding in the 1970s and 1980s was dominated by the study of folding intermediates [3,4]. Two alternative models prevailed, the first of which was the framework model [3–6] and the related diffusion–diffusion model [7], in which secondary structure is proposed to fold first, followed by docking of the pre-formed secondary structural units to yield the native, folded protein. The second was the hydrophobic collapse model [8–11], in which hydrophobic collapse drives compaction of the protein so that folding can take place in a confined volume, thereby narrowing the conformational search to the native state (Fig. 1).

## Classical evidence for mechanisms

The framework model gained support from studies mostly carried out in the 1980s on small, relatively stable, helical peptides [12–14]. Previously, it was assumed that secondary

structural segments were not stable enough to form in the absence of tertiary contacts [15]. When secondary structure is stable in isolation, it tends to adopt  $\alpha$ -helical and turn structures or, less frequently,  $\beta$ -structures in the form of  $\beta$ -hairpins. These structures might represent the starting-point for folding such that folding effectively proceeds from a denatured state with high secondary structure content. In practice, however, very strong conformational preferences are rare, and most peptides that form regular secondary structures in proteins are >90% disordered in small peptides containing those protein



**Fig. 1.** (a) Schematic view of protein folding showing the two classical, extreme models of protein folding: the hydrophobic collapse model and framework models. The nucleation-condensation model is a combination of the two models, and it can shift to either of the models with changes in the relative stability of the secondary or tertiary structures. (b) Examples of proteins following 'pure' forms of the models, as well as the shift or 'mixing' of the models. Proteins that fold via the framework model are entirely helical, reflecting the stability of local interactions and the difficulty in forming stable  $\beta$ -strands and the tertiary nature of  $\beta$ -sheets. By contrast, there are examples of  $\alpha$ ,  $\beta$ , and  $\alpha,\beta$ -mixed proteins that fold via the nucleation-condensation mechanism. We are not aware of any proteins that fold purely through nonspecific hydrophobic collapse. Instead, collapse is generally accompanied by secondary structure formation, which falls into the more concerted and coupled 'middle-of-the-road', nucleation-condensation mechanism. Proteins listed are discussed in the text aside from: protein A [48], tenascin [49], FKBP12 [50] and lysozyme [51]. Abbreviations: C12, chymotrypsin inhibitor 2; D, denatured state; FKBP12, FK506-binding protein; N, native state.

sequences, as observed initially by Epanand and Scheraga [15]. Consequently, hydrophobic and other interactions with the rest of the protein stabilize the unstable elements of secondary structure. This is clearly seen in series of fragments of chymotrypsin inhibitor 2 (CI2) of increasing length, in which the  $\alpha$ -helical and other secondary structural elements are not observed until residues that make long-range contacts (i.e. significantly separated in sequence) are present [16].

Support for the hydrophobic-collapse model came from early studies showing that the hydrophobic driving force provided by the expulsion of water from the burial of non-polar surfaces is substantial [9]. However, the idea that the conformational search is facilitated within a nonspecific hydrophobic globule presents a problem because an excess of interactions will hinder reorganization of both the polypeptide chain and side chains. The discussions that followed incorporated the molten globule intermediate [4,17], proposing that the secondary structure is formed in the process of collapsing, thus, edging closer to the middle-of-the-road in Fig. 1 (somewhere between the two existing models).

Another problem with the simple views of protein folding provided by the hydrophobic collapse and framework models that has become evident in the past decade is that the denatured state is rarely a random, unstructured coil in which side-chain interactions are fleeting and all amino acids behave independently. In general, very harsh conditions are necessary to obtain such disordered states. Instead, proteins generally adopt residual structure, which can be in the form of fluctuating secondary structure and dynamic side-chain interactions, particularly hydrophobic clusters [18,19]. This residual structure can be native-like or non-native. Residual structure can be just a product of the tendency of the chain to limit its solvent accessible surface area, or the structure can be beneficial in folding by helping to channel or bias the protein in its quest for the native state.

Experimental techniques for detecting structure during folding have improved with respect to both spatial and temporal resolution, and it has become clear that many proteins fold by almost running down the middle of the two extreme situations for folding (see Fig. 1); that is, secondary and tertiary structure formation tends to be coupled. Uversky and Fink [20] have provided support for this middle-of-the-road folding by analyzing data on the conformational properties of 41 native- and partially folded-states (i.e. proteins that populate stable equilibrium intermediates). They found a good correlation between the decrease in hydrodynamic volume and increase in secondary structure during folding. They found no evidence for either compact intermediates lacking secondary structure or highly ordered secondary structure in very expanded states.

### A new mechanism for folding: nucleation-condensation

The three classical mechanisms were thrown back into the 'melting pot' in the early 1990s by two events: (1) the discovery that proteins could fold by simple two-state kinetics, without the accumulation of folding intermediates [the archetype being CI2 (Fig. 2a)] [21]; and (2)

$\Phi$ -value analysis of the transition state (Box 1), which showed that secondary and tertiary structure are formed in parallel as CI2 undergoes a general collapse [22]. This work led to the nucleation-condensation (or nucleation-collapse) mechanism [23–25], which unites features of both the hydrophobic collapse and framework mechanisms. Nucleation-condensation invokes the formation of long range and other native hydrophobic interactions in the transition state to stabilize the otherwise weak secondary structure. As with most proteins, the transition state also forms stable folds owing to a combination of long-range tertiary interactions and secondary structure. Isolated elements of repeating secondary structure, such as  $\alpha$ -helices or  $\beta$ -hairpins, tend to have weak conformational preferences in the absence of the rest of the protein, and are stabilized by tertiary interactions that are made with the rest of the protein. The transition state resembles a distorted form of the native structure, with the least distorted part being loosely defined as the nucleus and the distortion tending to increase with increasing distance from the nucleus (reviewed in [26]).

Simplified lattice-model calculations, which depict proteins as beads on a string fixed to points in a 2D or 3D lattice, also support the idea of an extended but finite nucleus in the transition state [27,28].  $\Phi$ -values and general aspects of the folding process are well reproduced by funnel landscapes [29,30] and polymer physics models, such as free energy functionals, as developed by Wolynes and colleagues [31,32]. Thus, both high- and low-resolution theoretical studies are in good agreement regarding the general features of the nucleation-condensation model.

However, whether any or all of the classical or novel mechanisms occur in general, and whether there is an underlying unifying mechanism remains to be seen. The folding pathways of several small proteins have been analyzed at atomic resolution by both  $\Phi$ -value analysis and molecular dynamics (MD) simulation (reviewed in [33–34]; Box 2), and specific details and a general picture are emerging.

### Atomic level descriptions of folding

Molecular dynamics simulations combined with  $\Phi$ -value analysis and NMR studies are beginning to describe full folding–unfolding pathways at atomic resolution. Even if a protein appears to be following one of the models described above, or if an intermediate is not observed, there is always more happening below the surface. For example, at the atomic level, two-state folding is fundamentally impossible. This idea can be illustrated by the results of an atomistic MD simulation of the unfolding of CI2 (Fig. 2a) in water at 100°C in reverse time (i.e. viewed in the direction of folding) (Fig. 2b).

There is little residual structure in the denatured state of CI2, but some hydrophobic clusters near the center of the chain are evident, as well as fluctuating native helical structure (Fig. 2b). This MD-generated residual structure was confirmed by NMR studies [35]. Through fluctuation, gyrations and interactions within the protein and with solvent, the gross topology of the native state begins to emerge (e.g. 30 ns; Fig. 2b). Thus, although a protein might appear to fold in a two-state manner, inspection of

### Box 1. $\Phi$ -value analysis for detailed characterization of transition and intermediate states

Protein engineering techniques have had a profound effect on the field of protein folding. For example, researchers can now introduce spectroscopic probes into a protein, thus providing a means of monitoring the folding process. However, more comprehensive mapping of the folding process can be obtained through the systematic study of the energetic consequences of introducing mutations throughout the protein, and is termed the protein engineering method or  $\Phi$ -value analysis [a–c]. As the effects of the mutations are evaluated by both kinetic and equilibrium experiments, this method yields more information about transitory states relevant to the folding–unfolding process, such as the transition and intermediate states. The mutations act as probes, such that the structure at the site of mutation can be inferred from the energetics. More specifically, characterization of a structure of interest is based on a quantity  $\Phi$ , which is calculated from the following equation (Eqn 1),

$$\Phi = (\Delta G_{TS-D} - \Delta G'_{TS-D}) / (\Delta G_{N-D} - \Delta G'_{N-D}) = \Delta \Delta G_{TS-D} / \Delta \Delta G_{N-D} \quad (\text{Eqn 1})$$

where  $\Delta G_{TS-D}$  and  $\Delta G_{N-D}$  are the free energies of the transition state (this could also be an intermediate state) and the native state, respectively, relative to the denatured state for the wild-type protein (Fig. 1).

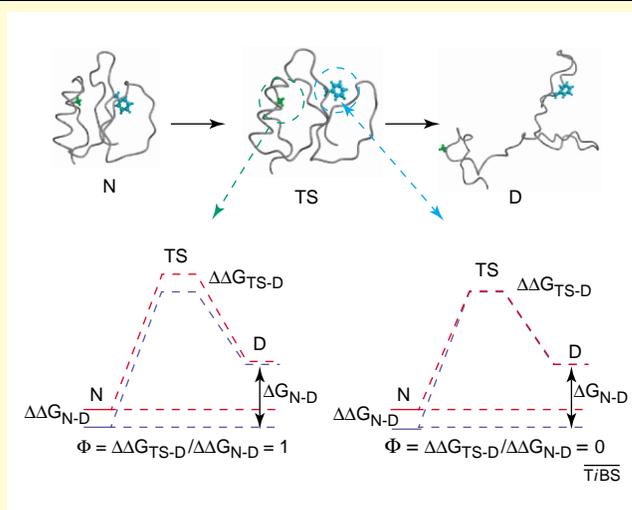
Corresponding terms for the mutant are indicated by a prime.  $\Delta \Delta G_{N-D}$  and  $\Delta \Delta G_{TS-D}$  are the destabilization energies of the native state and transition states of interest, respectively, caused by the mutation.

#### References

- Leatherbarrow, R.J. *et al.* (1985) Transition-state stabilization in the mechanism of tyrosyl-tRNA synthetase revealed by protein engineering. *Proc. Natl. Acad. Sci. U.S.A.* 82, 7840–7844
- Fersht, A.R. *et al.* (1986) Quantitative analysis of structure–activity relationships in engineered proteins by linear free-energy relationships. *Nature* 322, 284–286
- Fersht, A.R. *et al.* (1992) The folding of an enzyme I. Theory of protein engineering analysis of stability and pathway of protein folding. *J. Mol. Biol.* 224, 771–782
- Daggett, V. *et al.* (1996) Structure of the transition state for folding of a protein derived from experiment and simulation. *J. Mol. Biol.* 257, 430–440

snapshots of the protein shows that this is not the case at the atomic level (Fig. 2b). Nucleation sites are evident and collapse and condensation occur about these sites. However, the folding nucleus, which comprises residues distant in sequence and effectively spans the full sequence, does not come together until the transition state (Fig. 2b). Given the spacing of these residues, the nucleus necessarily reflects the consolidation of both secondary and tertiary structure. In addition, there is degeneracy built into the nucleus such that it can be better described as patches of the structure that coalesce in the transition state. The transition state identified from the simulation is in quantitative agreement with the experimental  $\Phi$ -values (Fig. 2c) [22,23,36–38]. It is important to note that all structures along the reaction pathways are ‘ensembles’, representing a statistical distribution around the canonical structure: the denatured state is a very diverse ensemble, the native state is a very tight ensemble, and the transition state is between the two, although generally closer to the native structure.

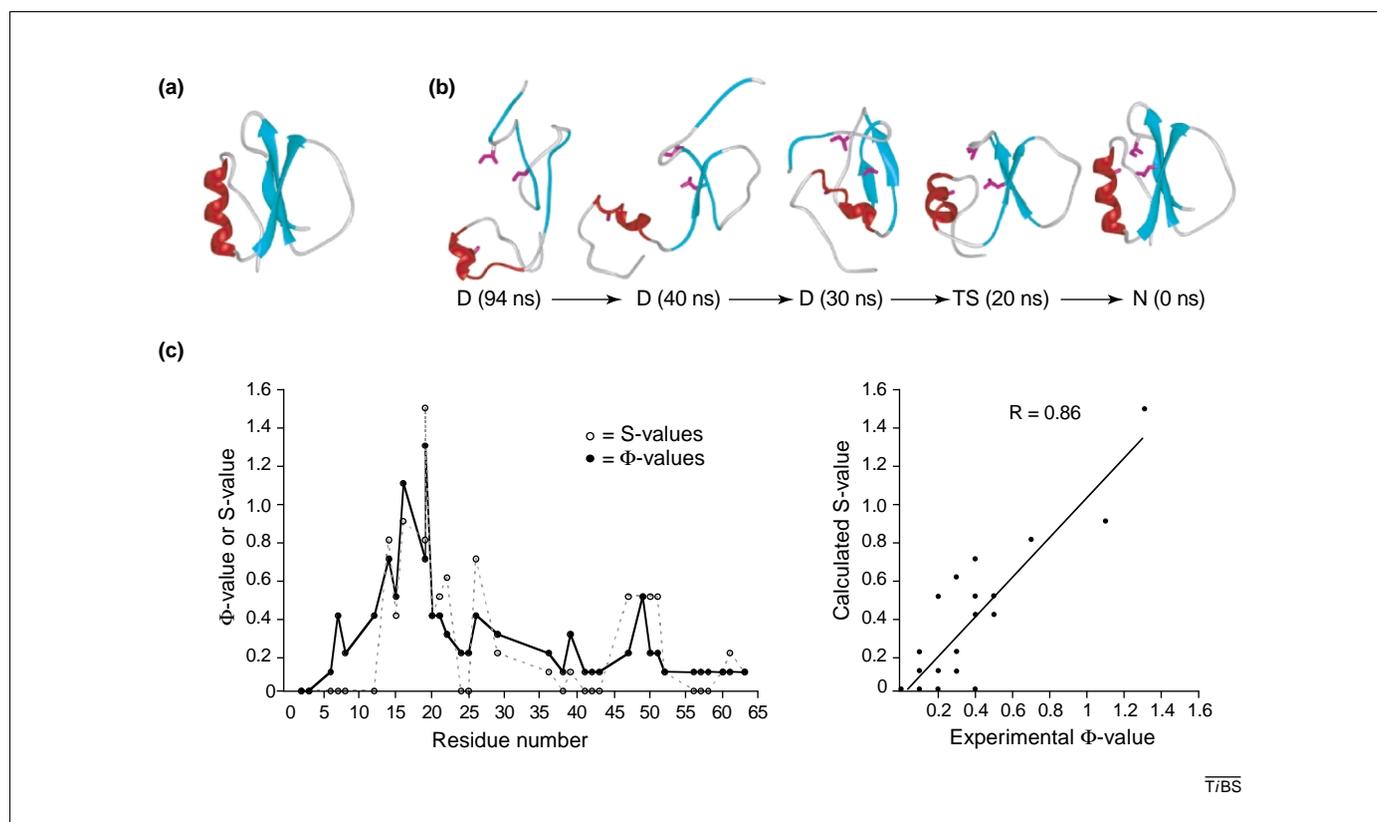
Interestingly, despite the variety of interactions occurring in the denatured state as the protein moves toward the native state, the gross behavior of the protein is quite two-state. In effect, experiment is blind to many of these



**Fig. 1.** In the transition state of unfolding, the structure of the protein at the site of mutation is the same as in the native state. The protein is immune to the effect of the mutation until after the major transition state, and the transition state is destabilized by exactly the same amount as the native state, that is,  $\Delta \Delta G_{TS-D} = \Delta \Delta G_{N-D}$ , and  $\Phi = 1$  (green residue). Conversely, a  $\Phi$  value of 0 means that the structure of the transition state at the site of mutation is the same as in the denatured state (cyan residue). Intermediate values represent structures that are partially unfolded in the transition state (or a mixture of folded and unfolded). The site of mutation has native-like extent of structure in the transition state when  $\Phi = 1$ . A residue-specific structure index (S-value) for transition state-ensembles generated by molecular dynamics is calculated for comparison with experimentation. S-values reflect the extent of packing interactions of the residue and its local secondary structure, and these values typically fall between 0 and 1 [d]. Abbreviations: D, denatured state; I, intermediate state; TS, transition state; N, native state.

other transient but crucial conformations, but that is not to say that they do not occur. As experimental techniques for protein folding have improved and are now able to monitor folding on faster time scales of ns– $\mu$ s, we have seen that proteins fold much faster than previously believed, closing the gap between the MD simulations (also ns– $\mu$ s) and experiment [39]. In addition, as the spatial resolution of experimental techniques improve and single-molecule measurements mature, so too will we see convergence of simulation and experiment.

Although the nucleation-condensation mechanism was introduced to help explain the behavior of apparently two-state folding proteins, it is not limited to small, single-domain proteins. This mechanism is also applicable to larger proteins by considering the behavior of the individual domains. The small, multi-domain protein, barnase, illustrates this phenomenon. Barnase contains two, semi-autonomous domains (Fig. 3a). Its folding–unfolding pathway has also been mapped in detail by combining experimental studies with simulation. Barnase contains significant residual structure in its denatured state as probed by MD, NMR and other experimental techniques [40–42]. This residual structure helps to set up loose, native topology within the denatured state (Fig. 3b). Even in the denatured state,



**Fig. 2.** The folding of chymotrypsin inhibitor 2 (CI2). (a) Crystal structure of CI2 [52]. (b) Snapshots, omitting the solvent, from the 100°C unfolding simulation shown in reverse time [53]. Residues important to the folding nucleus are displayed in magenta. Structures are labeled D, TS, or N, for denatured, transition, and native states, respectively. Time points for the structures are given in parentheses. (c) A comparison of experimental  $\Phi$ -values and calculated S-values for the transition state shown in (b) as a function of the site of mutation are shown on the left (see Box 1). A scatter-plot of the  $\Phi$ - and S-values is shown on the right. Protein images were made using UCSF MidasPlus [54].

its two domains remain semi-independent and are ‘pinched off’ from one another.

In the primary domain of barnase (see Fig. 3), which houses the main hydrophobic core, folding is nucleated by residual structure in the form of fluctuating native helical structure and hydrophobic clusters centered on the  $\beta(3-4)$  hairpin. Interestingly, the hairpin participates in the folding of the helix by actively aiding in its formation of secondary structure via side-chain interactions that help to pull the helix through dihedral angle transitions (Fig. 4). We have termed this ‘contact-assisted’ secondary structure formation. This type of mechanism is crucial in providing the link between secondary and tertiary structure for the more complicated processes that do not merely involve docking of pre-formed units of secondary structure.

In barnase, the hairpin and helix form a scaffold, providing a site for consolidation of the remaining  $\beta$ -strands (Fig. 3b). The resulting structure is loose and represents the major intermediate during the re-folding of barnase. The MD-generated intermediate shown in Fig. 3b is in quantitative agreement with the experimentally derived  $\Phi$ -values [43,44]. Further collapse and consolidation about this structure then follows in the transition state, bringing the residues involved in the folding nucleus into close proximity. As with the intermediate, the calculated and experimental  $\Phi$ -values are in good agreement (Fig. 3c) [45,46]. For the second helical domain, helix 2

contains considerable residual structure in the denatured state, and the short 3–10 helix docks upon it. Owing to the pinching off of this short loop, tertiary interactions are primed for the later collapse and refinement of packing interactions that occur in the transition state and the consolidation of the interface between the two domains.

Although barnase is small, it is reasonable to assume that other larger proteins will follow a similar pathway and pinch-off smaller domains to fold more locally. As such, it would appear that we have a unifying mechanism of protein folding. In the case of the framework model, it can be considered as an extension of nucleation-condensation, in which secondary structure is overly stable – particularly  $\alpha$ -helices (Fig. 1) – as with the engrailed homeodomain [39] (Fig. 5a). In this case, significant residual structure in the denatured state is propagated to the intermediate state (or, in effect, the denatured and intermediate states are equivalent) and the rate-determining transition state involves docking of these secondary structure elements and consolidation of packing interactions around these docking surfaces (Fig. 5b). Thus, if the secondary structure is sufficiently stable in the absence of significant tertiary interactions, folding can proceed in a more stepwise, hierarchical manner. In other members of the homeodomain superfamily, there is a shift from the diffusion–collision mechanism to nucleation-condensation as the intrinsic tendency to form stable

### Box 2. Molecular dynamics simulations of protein unfolding

Molecular dynamics simulations can be used to obtain atomic positions as a function of time, thereby providing a realistic and complete picture of protein motion. This technique requires a well-defined starting structure and a potential energy function whose parameters have been derived to reproduce structures and energy trends in various model systems. During the simulation, the atoms move due to their own kinetic energy and the forces exerted upon them by all other atoms. Specific interactions can then be monitored both geometrically and energetically to investigate structural transitions and the mechanism by which they occur. Given the importance of solvent in determining the conformational properties of proteins, it is imperative to include solvent in the simulation to model denaturation realistically (see Fig. 1). Because of severe sampling limitations under folding conditions with such high-resolution simulation methods, we typically focus on unfolding events (Fig. 1).

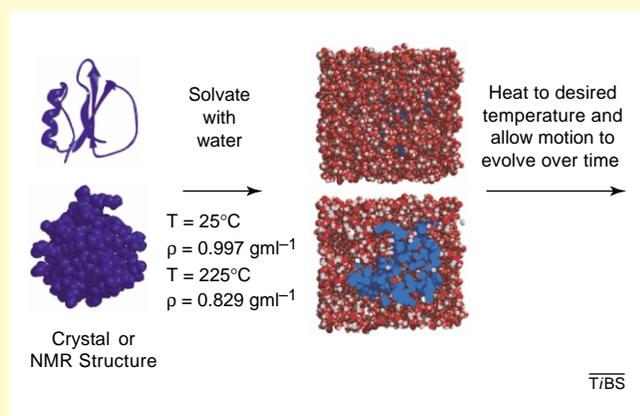


Fig. 1. Setting up a molecular dynamics simulation. The solvent box is at 225°C with the density set to the experiment value for this temperature.

There are numerous advantages to studying unfolding rather than folding. Simulations begin from a well-defined starting structure – a crystal or NMR structure – which improves the probability of sampling experimentally relevant regions of conformational space. It is not necessary to have molecular dynamics samples all of conformational space, as real proteins do not sample all possible conformations during

folding and unfolding. Another advantage to studying unfolding is that the full reaction coordinate, from the native to denatured states, can be explored. The temperatures used to unfold the protein are typically very high (225°C). However, it is worth noting that care must be taken to set the solvent density to the experimental value of the temperature of interest so that the excess pressure can be reduced and the water will remain a liquid up to 225°C (the water shown here is at 225°C, with the density set to the experimental value for this temperature). Also, we have shown that increasing the temperature merely accelerates the unfolding process, it does not change the overall pathway [b–d].

As with experiment, studying transition states presents problems for simulation studies. The transition state for folding–unfolding is an ensemble of high free-energy structures. Unfortunately, even if a reasonable unfolding pathway can be simulated with molecular dynamics, the calculation of free energies for such a complicated process is not possible. Instead, the structural properties, which are the strength of force field methods, are used to identify the transition state in a simulation. Using this approach, structural attributes of the transition state ensemble can be precisely delineated; however, there is no guarantee that the ensemble identified is the state of highest free energy. As the transition state is kinetically and thermodynamically unstable, the structure of the protein is expected to change rapidly once it passes the major transition state. Therefore, the major transition state in the simulations can be defined as the ensemble of structures populated immediately before the onset of a large structural change. The simplest method of identifying transition state regions is through a conformational cluster analysis [a].

### References

- a Li, A. and Daggett, V. (1994) Characterization of the transition state of protein unfolding by use of molecular dynamics: chymotrypsin inhibitor 2. *Proc. Natl. Acad. Sci. U.S.A.* 91, 10430–10434
- b Mayor, U. *et al.* (2000) Protein folding and unfolding in microseconds to nanoseconds by experiment and simulation. *Proc. Natl. Acad. Sci. U.S.A.* 97, 13518–13522
- c Ferguson *et al.* (2001) Using flexible loop mimetics to extend  $\Phi$ -value analysis to secondary structure interactions. *Proc. Natl. Acad. Sci. U.S.A.* 98, 13008–13013
- d Day, R. *et al.* (2002) Increasing temperature accelerates protein unfolding without changing the pathway of unfolding. *J. Mol. Biol.* 322, 189–203

secondary structure decreases (N.R. Guydosh *et al.*, unpublished results). Similarly, increasing hydrophobic content can favor formation of ‘molten globule’-like intermediates [47], possibly leading to a shift in the mechanism.

### Concluding remarks

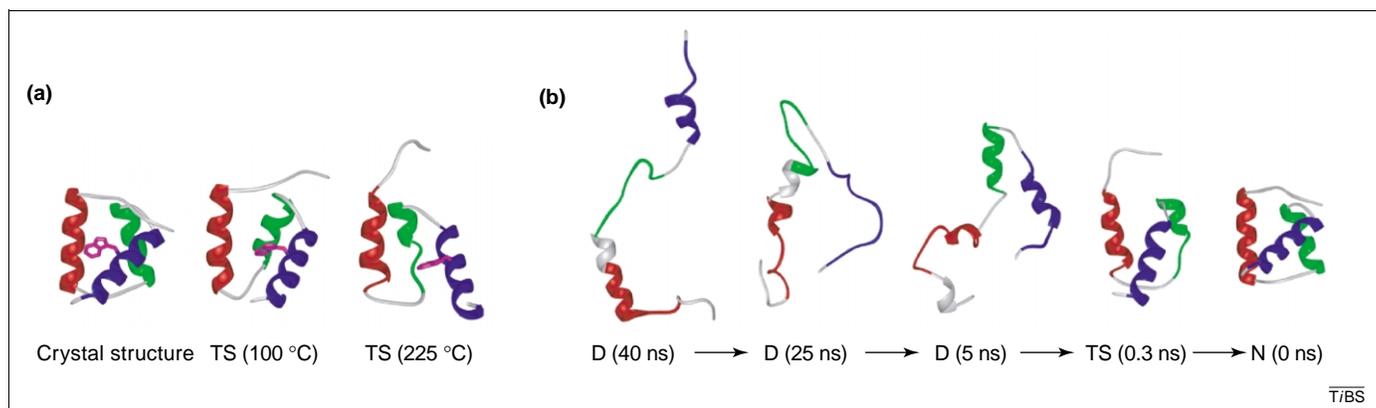
The framework and hydrophobic collapse mechanisms can be viewed as extremes of the nucleation-condensation mechanism: in a typical domain in which the intrinsic conformational preferences for secondary are weak, the formation of the transition state requires a considerable network of tertiary interactions to stabilize it. In domains that have unusually strong conformational preferences, such as the engrailed homeodomain, the folding appears to be hierarchical, and some proteins might fall into a trap of a molten globule if hydrophobic interactions are formed too rapidly and strongly. Accordingly, the nucleation-condensation mechanism appears to describe nicely the folding of proteins regardless of their size and the complexity of the folding pathway; that is, it is not limited to small single-domain, two-state folding proteins. In its original formulation, this model described the

folding of two-state proteins via concerted and cooperative secondary and tertiary structure formation about the transition state for folding. In the more general interpretation of the model described here, heightened stability of secondary structure can shift the mechanism to account not only for the formation of folding intermediates but also for hierarchical models of folding in general. In addition, through combined theoretical and experimental efforts, a novel mode of secondary structure formation via a tertiary-contact-assisted mechanism was discovered. This mechanism allows for generalization of the nucleation-condensation model and provides a detailed view of the coupling between secondary and tertiary structure. Whether this coupling is weak – favoring the framework model – or strong and occurring primarily within the transition state (as for single domain, two-state folding proteins) depends on the particular sequence and balance of forces.

### Acknowledgements

V.D. is grateful for financial support provided by the National Institutes of Health (GM 50789).





**Fig. 5.** The folding pathway of the engrailed homeodomain. (a) The crystal structure [56] and representative transition state structures from 100°C and 225°C simulations, showing how high temperature accelerates folding without changing the overall pathway of unfolding [39]. (b) Snapshots from the 225°C simulation are shown in reverse, and labeled by time and conformational state. The helices have strong helical propensities in isolation such that the rate-determining step involves docking of the helices. Abbreviations: D, denatured state; TS, transition state; N, native state. Protein images were made using MidasPlus (University of California, San Francisco, CA, USA) [54].

## References

- Levinthal, C. (1969) How to fold graciously. In *Mossbauer Spectroscopy in Biological Systems In Proceedings of a Meeting held at Allerton House* (Monticello, I.L., ed.), pp. 22–24, University of Illinois Press
- Wetlaufer, D.B. (1973) Nucleation, rapid folding, and globular intrachain regions in proteins. *Proc. Natl. Acad. Sci. U.S.A.* 70, 697–701
- Kim, P.S. and Baldwin, R.L. (1982) Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. *Annu. Rev. Biochem.* 51, 459–489
- Ptitsyn, O.B. (1987) Protein folding: hypotheses and experiments. *J. Protein Chem.* 6, 273–293
- Ptitsyn, O.B. and Rashin, A.A. (1975) A model of myoglobin self-organization. *Biophys. Chem.* 3, 1–20
- Kim, P.S. and Baldwin, R.L. (1990) Intermediates in the folding reactions of small proteins. *Annu. Rev. Biochem.* 59, 631–660
- Karplus, M. and Weaver, D.L. (1976) Protein folding dynamics. *Nature* 260, 404–406
- Schellman, J.A. (1955) Stability of hydrogen bonded peptide structures in aqueous solution. *Compt. Rend. Lab. Carlsberg. Ser. Chim.* 29, 230–259
- Kauzmann, W. (1959) Some factors in the interpretation of protein denaturation. *Adv. Protein Chem.* 14, 1–63
- Tanford, C. (1962) Contribution of hydrophobic interactions to the stability of globular confirmation of proteins. *J. Am. Chem. Soc.* 84, 4240
- Baldwin, R.L. (1989) How does protein folding get started? *Trends Biochem. Sci.* 14, 291–294
- Brown, J.E. and Klee, W.A. (1971) Helix–coil transition of the isolated amino terminus of ribonuclease. *Biochemistry* 10, 470–476
- Bierzynski, A. *et al.* (1982) A salt bridge stabilizes the helix formed by isolated C-peptide of RNase A. *Proc. Natl. Acad. Sci. U.S.A.* 79, 2470–2474
- Shoemaker, R. *et al.* (1985) Nature of the charged-group effect on the stability of the C-peptide helix. *Proc. Natl. Acad. Sci. U.S.A.* 82, 2349–2353
- Epand, R.M. and Scheraga, H.A. (1968) The influence of long-range interactions on the structure of myoglobin. *Biochemistry* 7, 2864–2872
- de Prat Gay, G. *et al.* (1995) Folding of a nascent polypeptide chain *in vitro*: cooperative formation of structure in a protein module. *Proc. Natl. Acad. Sci. U.S.A.* 92, 3683–3686
- Ptitsyn, O.B. (1995) Structures of folding intermediates. *Curr. Opin. Struct. Biol.* 5, 74–78
- Dobson, C.M. (1992) Unfolded proteins, compact states and molten globules. *Curr. Opin. Struct. Biol.* 2, 6–12
- Shortle, D. (1996) The denatured state (the other half of the folding equation) and its role in protein stability. *FASEB J.* 10, 27–34
- Uversky, V.N. and Fink, A.L. (2002) The chicken–egg scenario of protein folding revisited. *FEBS Lett.* 515, 79–83
- Jackson, S.E. and Fersht, A.R. (1991) Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. *Biochemistry* 30, 10428–10435
- Otzen, D.E. *et al.* (1994) Structure of the transition state for the folding/unfolding of the barley chymotrypsin inhibitor 2 and its implications for mechanisms of protein folding. *Proc. Natl. Acad. Sci. U.S.A.* 91, 10422–10425
- Itzhaki, L.S. *et al.* (1995) Detailed structure of the transition state for folding of chymotrypsin inhibitor 2 analyzed by protein engineering. *J. Mol. Biol.* 254, 260–288
- Fersht, A.R. (1995) Optimization of rates of protein folding: the nucleation-condensation mechanism and its implications. *Proc. Natl. Acad. Sci. U.S.A.* 92, 10869–10873
- Fersht, A.R. (1997) Nucleation mechanisms in protein folding. *Curr. Opin. Struct. Biol.* 7, 3–9
- Daggett, V. (2002) Molecular dynamics simulations of the protein unfolding/folding reaction. *Acc. Chem. Res.* 35, 422–429
- Abkevich, V.I. *et al.* (1994) Specific nucleus as the transition state for protein folding: evidence from the lattice model. *Biochemistry* 33, 10026–10036
- Klimov, D.K. and Thirumalai, D. (2001) Multiple protein folding nuclei and the transition state ensemble in two-state proteins. *Proteins* 43, 465
- Onuchic, J.N. *et al.* (1996) Protein folding funnels: the nature of the transition state ensemble. *Fold. Des.* 1, 441–450
- Clementi, C. *et al.* (2000) Topological and energetic factors: what determines the structural details of the transition state ensemble and ‘en route’ intermediates for protein folding? *J. Mol. Biol.* 298, 937–953
- Shoemaker, B.A. *et al.* (1997) Structural correlations in protein folding funnels. *Proc. Natl. Acad. Sci. U.S.A.* 94, 777–782
- Shoemaker, B.A. *et al.* (1999) Exploring structures in protein folding funnels with free energy functionals: the transition state ensemble. *J. Mol. Biol.* 287, 675–694
- Daggett, V. and Fersht, A.R. (2000) Transition states in protein folding. In *Mechanisms of Protein Folding*, 2nd edn, (Pain, R.H., ed.), pp. 175–211, Oxford University Press
- Fersht, A.R. and Daggett, V. (2002) Protein folding and unfolding at atomic resolution. *Cell* 108, 573–582
- Kazmirski, S.L. *et al.* (2001) Protein folding from a highly disordered denatured state: the folding pathway of chymotrypsin inhibitor 2 at atomic resolution. *Proc. Natl. Acad. Sci. U.S.A.* 98, 4349–4354
- Li, A. and Daggett, V. (1994) Characterization of the transition state of protein unfolding by use of molecular dynamics: chymotrypsin inhibitor 2. *Proc. Natl. Acad. Sci. U.S.A.* 91, 10430–10434
- Daggett, V. *et al.* (1996) Structure of the transition state for folding of a

- protein derived from experiment and simulation. *J. Mol. Biol.* 257, 430–440
- 38 Li, A. and Daggett, V. (1996) Identification and characterization of the unfolding transition state of chymotrypsin inhibitor 2 by molecular dynamics simulations. *J. Mol. Biol.* 257, 412–429
- 39 Mayor, U. *et al.* (2000) Protein folding and unfolding in microseconds to nanoseconds by experiment and simulation. *Proc. Natl. Acad. Sci. U.S.A.* 97, 13518–13522
- 40 Arcus, V.L. *et al.* (1995) A comparison of the pH, urea and temperature-denatured states of barnase by heteronuclear NMR: implications for the initiation of protein folding. *J. Mol. Biol.* 254, 305–321
- 41 Bond, C.J. *et al.* (1997) Characterization of residual structure in the thermally denatured state of barnase by simulation and experiment: description of the folding pathway. *Proc. Natl. Acad. Sci. U.S.A.* 94, 13409–13413
- 42 Wong, K.B. *et al.* (2000) Towards a complete description of the structural and dynamic properties of the denatured state of barnase and the role of residual structure in folding. *J. Mol. Biol.* 296, 1257–1282
- 43 Li, A. and Daggett, V. (1998) The unfolding of barnase: characterization of the major intermediate. *J. Mol. Biol.* 275, 677–694
- 44 Matouschek, A. *et al.* (1992) The folding of an enzyme IV. Structure of an intermediate in the refolding of barnase analyzed by a protein engineering procedure. *J. Mol. Biol.* 224, 819–835
- 45 Serrano, L. *et al.* (1992) The folding of an enzyme III. Structure of the transition state for unfolding of barnase analyzed by a protein engineering procedure. *J. Mol. Biol.* 224, 805–818
- 46 Daggett, V. *et al.* (1998) A combined molecular dynamics and  $\Phi$ -value analysis of structure–reactivity relationships in the transition state and unfolding pathway of barnase: the structural basis of hammond and anti-hammond effects. *J. Am. Chem. Soc.* 120, 12740–12754
- 47 Uversky, V.N. (2002) Cracking the folding code. Why do some proteins adopt partially folded conformations, whereas other don't? *FEBS Lett.* 514, 181–183
- 48 Myers, J.K. and Oas, T.G. (2001) Preorganized secondary structure as an important determinant of fast protein folding. *Nat. Struct. Biol.* 8, 552–558
- 49 Clarke, J. *et al.* (1997) Folding and stability of a fibronectin type III domain of human tenascin. *J. Mol. Biol.* 270, 771–778
- 50 Fulton, K.F. *et al.* (1999) Mapping the interactions present in the transition state for folding/unfolding of FKBP12. *J. Mol. Biol.* 291, 445–461
- 51 Matagne, A. *et al.* (1997) Fast and slow tracks in lysozyme folding: insight into the role of domains in the folding process. *J. Mol. Biol.* 267, 1068–1074
- 52 Harpaz, Y. *et al.* (1994) Direct observation of a better hydration at the N-Terminus of an  $\alpha$ -helix with glycine rather than alanine as the N-cap residue. *Proc. Natl. Acad. Sci. U.S.A.* 91, 3–15
- 53 Day, R. *et al.* (2002) Increasing temperature accelerates protein unfolding without changing the pathway of unfolding. *J. Mol. Biol.* 322, 189–203
- 54 Ferrin, T.E. *et al.* (1988) The MIDAS display system. *J. Mol. Graph.* 6, 13–27
- 55 Bycroft, M. *et al.* (1991) Determination of the three-dimensional structure of barnase using nuclear magnetic resonance spectroscopy. *Biochemistry* 30, 8697–8701
- 56 Clarke, N.D. *et al.* (1994) Structural studies of the engrailed homeodomain. *Protein Sci.* 3, 1779–1787

## Articles of interest in other journals

### Elevated levels of GluR1 in the midbrain: a trigger for sensitization to drugs of abuse?

W.A. Carlezon, Jr and E.J. Nestler, *Trends in Neurosciences* 25, 610–615

### Hyperacidification in cystic fibrosis: links with lung disease and new prospects for treatment

J. Poschet *et al.*, *Trends in Molecular Medicine* 8, 512–519

### Drugs targeting functional bowel disorders: insights from animal studies

G.J. Sanger and G.A. Hicks, *Current Opinion in Pharmacology* 2, 678–683

### Drug targeting to specific vascular sites

E. Ruoslahti, *Drug Discovery Today* 7, 1138–1143

### Therapy of autoimmune diseases: clinical trials and new biologics

G.T. Nepom, *Current Opinion in Immunology* 14, 812–815

### Botulinum and tetanus neurotoxins: structure, function and therapeutic utility

K. Turton *et al.*, *Trends in Biochemical Sciences* 27, 552–558

### Proteomics: drug target discovery on an industrial scale

T.E. Ryan and S.D. Patterson, *Trends in Biotechnology* 20, S45–S51

### Somatostatin analogs in the diagnosis and treatment of cancer

S.W.J. Lamberts *et al.*, *Trends in Endocrinology and Metabolism* 13, 451–457

### Growth factor treatment of demyelinating disease: at last, a leap into the light

R.M. Ransohoff *et al.*, *Trends in Immunology* 23, 512–516

### Private prescription: rat of the month

R.C. Rowe, *Drug Discovery Today* 7, 1118–1119