JPP JOURNAL OF PROTEINS AND PROTEOMICS
3(2), July-December 2012, pp. 85-93

RESEARCH EDUCATION

MODELS OF PROTEIN FOLDING

Unnati Ahluwalia#, Nidhi Katyal# and Shashank Deep*

Department of Chemistry, Indian Institute of Technology, New Delhi 110016, India
#Equal contribution

Abstract: In an attempt to explore the understanding of protein folding mechanism, various models have been proposed in the literature. Advances in recent experimental and computational techniques rationalized our understanding on some of the fundamental features of the protein folding pathways. The goal of this review is to revisit the various models and outline the essential aspects of the folding reaction.

Introduction

Protein folding is the process by which a newly synthesized unfolded protein molecule folds into its unique, three-dimensional conformation. Understanding protein folding is important since it will help in prediction of the native structure of a protein from the knowledge of its primary structure. However, protein folding is a very complex process and the molecular mechanisms responsible for protein assembly are of the most elemental open question in biochemistry. This review outlines various important hypothesis/ models proposed by researchers to understand the folding problem.

Levinthal Paradox - presence of intermediate(s)

In 1961, Christian B. Anfinsen (Anfinsen *et al.*, 1961) and his colleagues were able to show that a protein takes its specific shape based on the "directions" encoded in the sequence of amino acids. Anfinsen showed that the folding of RNase A proceeds spontaneously downhill to the lowest free-energy polypeptide conformation, which is its native state. Next important development took place when Cyrus Levinthal made the argument that "there are too many possible conformations for the unfolded protein to find the native state in conformational space by random searching"(Levinthal, 1968). The fact that many naturally-occurring proteins fold reliably and quickly to their native state (time scale of micro-milli second) despite the very large number of possible configurations has come to be known as Levinthal's Paradox. Levinthal concluded that random searches are not an effective way of finding the correct state of a folded protein.

Thus, researchers associated protein folding process to two mutually exclusive goals – achieving global minimum and doing so quickly (Dill and Chan, 1997). These two mutually exclusive goals are called thermodynamic and kinetic control. Thermodynamic control means that a protein reaches its global minimum in energy during folding and that the folding is independent of pathway taken to achieve this minimum. Kinetic controls means that folding happens quickly (on biological time scale) and thus protein may get stuck in local minimum over physiological time scale. Kinetic control is path dependent and final structure could be different depending on the denaturing conditions from which folding was initiated.

Levinthal's work motivated scientists to look for specific folding intermediates. A consensus

Corresponding Author: **Shashank Deep**
E-mail: sdeep@chemistry.iitd.ac.in

Received: November 26, 2012
Accepted: November 28, 2012
Published: November 30, 2012

emerged that if we could observe intermediate states along the folding pathway, we could learn how an unfolded protein finds its native structure from the large number of possible conformations. In this context, most notable work was carried out by the research groups of Baldwin (Kim and Baldwin, 1982, 1990) and Creighton (Creighton, 1986). Creighton and co-workers looked out for folding intermediates that could be trapped by covalent disulfide bond formation. Baldwin and co-workers employed a range of spectroscopic methods to observe folding in real time. They were also the first one to use the hydrogen exchange methods for exploring folding intermediates (Schmid and Baldwin, 1979).

Models of Folding

Different models have been established to explain the surprisingly high speed of protein folding, compared with a random sampling mechanism, e.g. the framework model, the nucleation-growth mechanism, the diffusion-collision mechanism, the hydrophobic collapse model and the energy landscape (folding funnel) model.

Classical view: Folding Pathways

Sequential Protein Folding Model

This model, also known as the framework or hierarchic model, was proposed in 1973 by Ptitsyn (Ptitsyn, 1973). This suggests that the folding starts with the formation of secondary structure elements, which then interact to form a more advanced folding intermediate. The folding process ends with the specific packing of the side chains. Thus each step forward in the folding process stabilizes the major structural elements formed at the previous stage, suggesting the existence of several folding intermediates. It is assumed that local elements of native secondary structure could be formed independent of tertiary structure. This hypothesis was boosted when it was revealed through the help of newer ultra-fast experiments that folding/ refolding of the individual elements of the secondary structure occurs on nanosecond time scale and thus can occur far before the main folding event (Jeng *et al.*, 1990; Matouschek *et al.*, 1989; Serrano *et al.*, 1992). But then how to decipher the fast folding kinetics of protein is still an open question. Further, the process of hierarchic condensation has not been verified experimentally for large number of proteins since the intermediate states of folding are generally very unstable (Go, 1984). In addition, whether the secondary structure elements are stable in isolation (Blanco *et al.*, 1998; Munoz and Serrano, 1996) or need some stabilizing tertiary contacts is debatable (De Prat Gay *et al.*, 1995). In nutshell, in this model local interactions dominate and guide the formation of secondary structural elements, followed by random diffusion collision of these local elements of the secondary structure until stable native tertiary contacts are made to give the final structure.

Diffusion Collision Model

This model was suggested in 1976 by Karplus and Weaver as a model for the process of protein folding (Karplus and Weaver, 1976). Infact framework model can be considered as a limiting case of this model (Karplus and Weaver, 1994). The primary assumption of the model is that protein is made up of several unstable quasiparticles called microdomains. These microdomains are simply portions of nascent secondary structure or hydrophobic clusters. Each of them is small and therefore all the conformational alternatives in this domain can be searched very rapidly as compared with the time scale of the entire folding process. These microdomains are less stable and to gain stability they diffuse, collide and finally coalesce. Two units coalesce forming a slightly more stable entity which in turn collides with third and so on. This way stability gets enhanced at each subsequent step and so does the rate of the process, i.e. diffusion collision becomes more probable with increase in stability of secondary structure. Each step leads to conformation closer to native structure and finally leading to formation of tertiary structure. Coalescence of microdomains not necessarily adjacent in the sequence might occur.

Short range attractive forces govern the diffusion motion rendering stability to the coalesced microdomains. Not every collision leads to coalition. The balance between hydrophobic attraction and van der waals repulsion is the deciding factor. This requires that

segments have correct secondary structure (including side chain orientation) else a significant repulsive barrier would be expected. However much of the information is still needs to be resolved. Model does not exclude the formation of non native intermediates (example in folding of proteins with disulphide bonds) (Karplus and Weaver, 1994).

During the diffusion process, if microdomains were considered to be flexible, various problems due to entanglement might arise and so these diffusing units are assumed to be rigid. Empirical correlations tend to support this view (Chou and Fasman, 1974; Crawford *et al*., 1973; Karplus and Weaver, 1976; Lewis *et al*., 1971). The overall dynamics of folding is governed by a set of diffusion equations that describe the relative motion of the microdomains and their aggregates and by boundary conditions which couple the diffusion equations through the collision and coalescence of the microdomains.

Hydrophobic Collapse Model

Hydrophobic collapse is a hypothesis proposed for the folding of proteins based on the observation that protein's native states often contain a hydrophobic core of nonpolar amino acid side chains in the protein's interior (Go, 1984). The energetic stabilization of folding intermediate is believed to be due to the sequestration of the hydrophobic side chains from the surrounding water. This collapsed intermediate is also referred to as a molten globule and corresponds to a partially folded state. Whether its energy is lower or higher than the denatured state is an interplay of balance between entropy and enthalpy terms. If the entropic stabilization (due to numerous available conformational fluctuations) outweighs the enthapic destabilization (due to poor packing of atoms in the interior of the globule), the globular state would be observed as an equilibrium state. It can be kinetically observed under strong refolding conditions in which its energy is lower than that of the denatured state but higher than that of native. On the other hand under weak refolding conditions when the state is unstable against both the unfolded and native states, it may not be observed even kinetically.

The hypothesis predicts that the hydrophobic collapse is relatively early event in the folding pathway (Sinha and Udgaonkar, 2009), occurring before the formation of secondary structures and native contacts present in the fully folded tertiary structure. This is consistent with the various lattice model simulations which shed light on the compactness of unfolded state in native conditions (Dobson *et al*., 1998; Sali *et al*., 1994). In addition, observation of a burst- phase loss in the signals of different spectroscopic probes in ms measurements is indeed suggestive of initial collapse in protein folding reactions (Agashe *et al*., 1995; Elove *et al*., 1992; Houry *et al*., 1996; Jennings and Wright, 1993; Khorasanizadeh *et al*., 1996; Park *et al*., 1997; Raschke and Marqusee, 1997). Further evidence in its favor includes a significant enhancement of the ANS fluorescence in the < 1ms time domain (Agashe *et al*., 1995; Engelhard and Evans, 1995; Mann and Matthews, 1993). However, it was observed through stopped flow SAXS measurement that acyl-phosphate and Ub variant F45W do not appear to undergo initial collapse reaction (Jacob *et al*., 2004). To reconcile the apparently conflicting evidence it was suggested (Sinha and Udgaonkar, 2009) that fast chain collapse may remain undetected in case of many proteins because of the nature of the probes utilized. There may also be the possibility of concurrent structure formation which may slow down the collapse reaction. On an average it can be assumed that folding begins with the clustering of hydrophobic amino acids via hydrophobic interactions. Now since these interactions are believed to be largely non-specific, it is expected that "a nonspecific globule" is formed first. It was postulated (Go, 1984) that transformation occurs from non specific globule to specific native globular structure either via random motion or by transition into the unfolded state followed by refolding into another non specific globule. Several experimental artifacts suggests initial collapse being non specific in case of RNase A(Qi *et al*., 1998; Sosnick *et al*., 1997).

It is difficult to discern whether hydrophobic collapse triggers secondary structure formation or whether the two occurs concomitantly or on different timescales. In an attempt to resolve the question several studies have been carried out on different proteins. The molten globule form of

apomyoglobin (Eliezer *et al.*, 1998) and bovine alpha lactalbumin (Balbach *et al.*, 1997) are both compact and structured. For RNase A and BBL (Sadqi *et al.*, 2003; Welker *et al.*, 2004), a fast non specific collapse precede structure formation. For CspB (Pradeep and Udgaonkar, 2002), cytochrome c (Akiyama *et al.*, 2002) and monellin (Kimura *et al.*, 2008; Kimura *et al.*, 2005) significant structure develops only after a fast collapse reaction. The kinetic data suggests that small energy barrier (~k_BT) is encountered by collapsing polypeptide chain and that the structure formation may be highly non cooperative and occurs along multiple tracks. This also accounts for hydrophobic collapse being the fast early event in protein folding.

Nucleation Condensation Model ("collapse around diffuse nucleus")

It is based on the assumption that protein folding is similar to the crystallization process, and that the limiting step in the folding process is the nucleus formation (Fersht, 1997), which is followed by the rapid propagation of structure. It is an amalgam of both hydrophobic collapse and framework mechanism. This model suggests that the folding starts with the formation of a more diffused nucleus. Unlike the classical nucleation wherein strong localized nucleus develops, the newer aspect of the model emphasizes on formation of weak local nucleus which is further stabilized by long range interactions leading to large extended nucleus. The building up of the nucleus is not initiated until the transition state is reached. This implies minimal intermediate formation thereby rendering it a suitable model in circumventing the Levinthal Paradox. But it would be incorrect to associate the folding of a protein with two state kinetics to the nucleation mechanism since hierarchical mechanism with a high energy intermediate would show similar kinetics. The key feature of nucleation condensation model is the concurrent formation of nucleus and secondary and tertiary structures. As the nucleus consolidates its local and long range interactions, the increase in stability takes place so rapidly that nucleus is not yet fully formed in the transition state (Itzhaki *et al.*, 1995). The formation of the nucleus is thus coupled with the condensation. Larger proteins can be thought to be made up of modules which can separately fold by nucleation condensation and then dock or fold by nucleation (Fersht, 1997).

We highlight a selection of recent examples in the form of a table (Table 1) of some of the proteins that undergoes folding via the above described models.

Recent studies (Daggett and Fersht, 2003) have revealed that some proteins unite features of various models. They can incorporate additional features from other models. In other words, the various competing pathways need not be mutually exclusive. The question of whether any underlying unifying mechanism governs the folding in general has been a contentious one. Some studies report that nucleation condensation mechanism appears to nicely describe the folding of proteins regardless of their size and the complexity of the folding pathway (Nolting and Agard, 2008). Finding has led to an idea that protein folds by multiple parallel pathways prior to nucleation. Nucleus is formed by zipper like process. This is then accompanied by hydrophobic collapse but concurrent formation of secondary and tertiary structures (in accordance with nucleation condensation mechanism). In the transition state where it is assumed that consolidation of local and long range interactions occur, it is observed that secondary structure forming residues are higher in number (supporting framework like features of nucleation).

New View: Energy Landscape and Folding Funnel

The energy landscape theory was first proposed by Joseph Bryngelson (Bryngelson et al., 1995) and José N. Onuchic (Onuchic et al., 1997). This landscape describes the dependence of the free energy on all the coordinates determining the protein conformation. The y-axis of the landscape represents the internal free energy of a given polypeptide configuration whereas lateral axes represent the conformational coordinates. Internal free energy includes the hydrogen bonds energy, ion pairs energy, torsion angle energy, hydrophobic and solvation free energy. The lateral axes have a large number of dimensionality reflecting the many degrees of

Table 1
Proposed folding models for different proteins

Protein	*Technique of Study*	*Mechanism*
FSD10 (Lee *et al*., 2012)	ADMD simulations	Hydrophobic Collapse followed by concurrent secondary and tertiary structure formation.
C-terminal domain of the Fas-associated death domain (Fadd-DD) (Greene *et al*., 2012)	quenched-flow hydrogen-deuterium exchange (NMR)	Hydrophobic Collapse. Secondary structure is largely concomitant with the hydrophobic collapse.
TC5b (Mok *et al*., 2007)	Photo CIDNP NMR	A pre-existing, hydrophobically collapsed conformation with both native-like and non-native interactions
CspB, CTL9, IM9, SpectrinR17, ubiquitin, SpectrinR16, apo-Azurin, FKBP12,IM7 (Nolting and Agard, 2008)	Φ- value analysis	Nucleation Condensation-almost concurrent build-up of secondary and tertiary structure contacts
Chymotrypsin Inhibitor 2 (Itzhaki *et al*., 1995)	Protein Engineering Methods	Nucleation Condensation
Ribonuclease A (Lustig and Fink, 1992)	CD, fluorescence, absorbance	Initial hydrophobic collapse, concurrent with secondary structure formation, followed by much slower rearrangement to the native tertiary structure
Cytochrome c (Gianni *et al*., 2003)	Stopped Flow experiments	Parallel Folding Pathway
Barnase (Fersht, 1993)	NMR	Framework
Engrailed homeodomain (Daggett and Fersht, 2003)	Φ-value analysis, MD simulations	Diffusion Collision
$\lambda_{6\text{-}85}$ repressor fragment (Myers and Oas, 1999)	Φ-value analysis, NMR	Diffusion Collision
C Myb transforming protein (Gilmanshin *et al*., 1997)	Brosnsted plot and MD simulations	Mixed Framework/Nucleation Condensation
hTRF1 (Daggett and Fersht, 2003)	Φ-value analysis, MD simulations	Nucleation Condensation
Myoglobin (Gilmanshin *et al*., 1997), Alpha lactalbumin (Arai and Kuwajima, 1996), barstar (Agashe *et al*., 1995), staphylococcal nuclease (Vidugiris *et al*., 1995)	Φ-value analysis, MD, MC simulations	Partial hydrophobic Collapse
Bc -Csp cold shock protein (Magg and Schmid, 2004)	FRET kinetics	Hydrophobic collpase
Monellin (Kimura *et al*., 2008)	Time resolved -IR	Hydrophobic colplaspe
BBL (peripheral subunit binding domain) (Welker *et al*., 2004)	Ultrarapid double-jump laser-induced temperature jump, kinetics Fluorescence	Hydrophobic collapse
B domain of protein A (Islam *et al*., 2002)	using stopped-flow CD and NMR hydrogen–deuterium exchange pulse labeling	diffusion–collision model
B1 domain of streptococcal protein G (Best and Hummer, 2011)	MD simulations	Diffusion Collision
Acylphosphatase (Muscle and common type) (van Nuland *et al*., 1998)	Fluorescence, CD kinetics	Nucleation Condensation
BphC enzyme (Zhou *et al*., 2004)	MD simulation	Hydrophobic Collapse

freedom available to a polypeptide chain. Each conformation is represented by a point on the energy landscape (Dill and Chan, 1997). High energy conformation looks like hill, whereas low energy conformation looks like a valley on the energy landscape. Free energy is high for large number of conformational states whereas it is small for small number of conformations. Therefore, this energetic surface is often called the "energy funnel".

The energy landscape theory states that folding of a protein does not follow a singular, specific pathway but occurs through routes down a folding funnel more like rain flowing down a funnel. On the highest energy level, proteins are in disordered state. As the proteins fold more into the organized, native-like conformation, they shift to a lower energy phase. At the end of the folding, proteins find their energy minima which correspond to their correctly packed native conformations with a unique set of α-helical and β-sheet motifs. The thermodynamic requirement for the energy minimum makes the protein folding highly efficient and very rapid. The cross section and topology of this folding funnel is unique for a specific polypeptide sequence under a particular set of conditions. However, the shape of the funnel may not be same for different polypeptide chain (Dill and Chan, 1997). A smooth funnel is expected for fast folders. A moat shaped energy landscape suggests the existence of an obligatory intermediate during folding. When folding is dominated by diffusional conformational search, a golf-course type energy landscape is expected. If the cooperativity of protein is not high, more than one intermediate may exist in the folding pathway. This will result into appearance of several local energy minima, local energetic traps, thus resulting into a rugged energy landscape. For some proteins, the kinetic traps are very deep and polypeptide chain will fold to some state which is energetically similar to the native state.

Several proteins have been suggested to fold via parallel routes, such as lysozyme (Kiefhaber, 1995), c-type cytochromes (Gianni *et al.*, 2003) and 27th immunoglobulin domain of titin (Wright *et al.*, 2003). It has been observed during experimental (Gianni *et al.*, 2007) and computational studies (Wallin *et al.*, 2007) that proteins do switch their preferred folding pathways depending on the environmental conditions. Molecular dynamic simulations have supported the existence of parallel folding pathways for a variety of small protein domains (Lazaridis and Karplus, 1997); (Borreguero *et al.*, 2004); (Rao *et al.*, 2005); (Juraszek and Bolhuis, 2006); (Lam *et al.*, 2007). The different transition states has also been observed for homologous proteins (McCallister *et al.*, 2000) and engineered circular permutants (Lindberg *et al.*, 2002); (Hubner *et al.*, 2006), supporting the existence of parallel folding pathways.

Old versus New view of Protein Folding

The "classical view" of protein folding is that the polypeptide searches for the native state through the enormous conformational space while flowing through predetermined pathways defined by discrete intermediates (Kim and Baldwin, 1990), (Baldwin, 1999), (Levinthal, 1968; Matthews, 1993). The "new view" of protein folding ((Bryngelson *et al.*, 1995), (Baldwin, 1995) (Dill and Chan, 1997) (Pande *et al.*, 1998) (Onuchic *et al.*, 2000) suggest that predefined pathways with compulsory intermediates simply do not exist. The downhill nature of the folding ensures that folding proceeds one amino acid at a time randomly. When discrete intermediate structures do become significantly populated, this could be inferred that the folding chain has accidentally wandered into some kinetic traps in the landscape, which only slows rather than promotes the folding process.

Computer simulation studies by Zwanzig group. has shown that, contrary to the Levinthal speculation, the search through conformational space might still be random if it is energetically biased to be downhill (Zwanzig *et al.*, 1992). No specific pathway would then be necessary. Continuing theoretical investigations have led to the picture of multiple energetically downhill paths through a funnel-shaped folding energy landscape (Bryngelson *et al.*, 1995; Leopold *et al.*, 1992; Plotkin and Onuchic, 2002a, b; Wolynes *et al.*, 1995). The landscape has been explored in many theoretical and experimental efforts, which have generally endorsed the multi-pathway view.

Summary

This review has discussed the various folding models proposed by the researchers. A lot of advancement has been made which explains how a poplypeptide chain folds. However, much needs to be done to understand what drives the folding.

Acknowledgement

This work is generously supported by the Council for Scientific and Industrial Research (CSIR), Government of India through grants to S.D. N. K. would like to thank Indian Institute of Technology, Delhi for the Fellowship.

References

Agashe, V. R., Shastry, M. C., and Udgaonkar, J. B. (1995). Initial hydrophobic collapse in the folding of barstar. Nature *377*, 754-757.

Akiyama, S., Takahashi, S., Kimura, T., Ishimori, K., Morishima, I., Nishikawa, Y., and Fujisawa, T. (2002). Conformational landscape of cytochrome c folding studied by microsecond-resolved small-angle x-ray scattering. Proc Natl Acad Sci U S A *99*, 1329-1334.

Anfinsen, C.B., Haber, E., Sela, M., and White, F.H., Jr. (1961). The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. Proc Natl Acad Sci U S A *47*, 1309-1314.

Arai, M., and Kuwajima, K. (1996). Rapid formation of a molten globule intermediate in refolding of alpha-lactalbumin. Fold Des *1*, 275-287.

Balbach, J., Forge, V., Lau, W.S., Jones, J.A., van Nuland, N.A., and Dobson, C.M. (1997). Detection of residue contacts in a protein folding intermediate. Proc Natl Acad Sci U S A *94*, 7182-7185.

Baldwin, R.L. (1995). The nature of protein folding pathways: the classical versus the new view. J Biomol NMR *5*, 103-109.

Baldwin, R.L. (1999). Protein folding from 1961 to 1982. Nat Struct Biol *6*, 814-817.

Best, R.B., and Hummer, G. (2011). Diffusion models of protein folding. Phys Chem Chem Phys *13*, 16902-16911.

Blanco, F., Ramirez-Alvarado, M., and Serrano, L. (1998). Formation and stability of beta-hairpin structures in polypeptides. Curr Opin Struct Biol *8*, 107-111.

Borreguero, J.M., Ding, F., Buldyrev, S.V., Stanley, H.E., and Dokholyan, N.V. (2004). Multiple folding pathways of the SH3 domain. Biophys J *87*, 521-533.

Bryngelson, J.D., Onuchic, J.N., Socci, N.D., and Wolynes, P.G. (1995). Funnels, pathways, and the energy landscape of protein folding: a synthesis. Proteins *21*, 167-195.

Chou, P.Y., and Fasman, G.D. (1974). Prediction of protein conformation. Biochemistry *13*, 222-245.

Crawford, J.L., Lipscomb, W.N., and Schellman, C.G. (1973). The reverse turn as a polypeptide conformation in globular proteins. Proc Natl Acad Sci U S A *70*, 538-542.

Creighton, T.E. (1986). Disulfide bonds as probes of protein folding pathways. Methods Enzymol *131*, 83-106.

Daggett, V., and Fersht, A.R. (2003). Is there a unifying mechanism for protein folding? Trends Biochem Sci *28*, 18-25.

De Prat Gay, G., Ruiz-Sanz, J., Neira, J.L., Itzhaki, L.S., and Fersht, A.R. (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.

Dill, K.A., and Chan, H.S. (1997). From Levinthal to pathways to funnels. Nat Struct Biol *4*, 10-19.

Dobson, C.M., Sali, A., and Karplus, M. (1998). Protein folding: A perspective from theory and experiment. Angew Chem Int Edit *37*, 868-893.

Eliezer, D., Yao, J., Dyson, H.J., and Wright, P.E. (1998). Structural and dynamic characterization of partially folded states of apomyoglobin and implications for protein folding. Nat Struct Biol *5*, 148-155.

Elove, G.A., Chaffotte, A.F., Roder, H., and Goldberg, M.E. (1992). Early steps in cytochrome c folding probed by time-resolved circular dichroism and fluorescence spectroscopy. Biochemistry *31*, 6876-6883.

Engelhard, M., and Evans, P.A. (1995). Kinetics of interaction of partially folded proteins with a hydrophobic dye: evidence that molten globule character is maximal in early folding intermediates. Protein Sci *4*, 1553-1562.

Fersht, A.R. (1993). The sixth Datta Lecture. Protein folding and stability: the pathway of folding of barnase. FEBS Lett *325*, 5-16.

Fersht, A.R. (1997). Nucleation mechanisms in protein folding. Curr Opin Struct Biol *7*, 3-9.

Gianni, S., Brunori, M., and Travaglini-Allocatelli, C. (2007). Plasticity of the protein folding landscape: switching between on- and off-pathway intermediates. Arch Biochem Biophys *466*, 172-176.

Gianni, S., Travaglini-Allocatelli, C., Cutruzzola, F., Brunori, M., Shastry, M.C., and Roder, H. (2003). Parallel pathways in cytochrome c(551) folding. J Mol Biol *330*, 1145-1152.

Gilmanshin, R., Dyer, R.B., and Callender, R.H. (1997). Structural heterogeneity of the various forms of apomyoglobin: implications for protein folding. Protein Sci *6*, 2134-2142.

Go, N. (1984). The consistency principle in protein structure and pathways of folding. Adv Biophys *18*, 149-164.

Greene, L.H., Li, H., Zhong, J., Zhao, G., and Wilson, K. (2012). Folding of an all-helical Greek-key protein monitored by quenched-flow hydrogen-deuterium exchange and NMR spectroscopy. Eur Biophys J *41*, 41-51.

Houry, W.A., Rothwarf, D.M., and Scheraga, H.A. (1996). Circular dichroism evidence for the presence of burst-phase intermediates on the conformational folding pathway of ribonuclease A. Biochemistry *35*, 10125-10133.

Hubner, I.A., Lindberg, M., Haglund, E., Oliveberg, M., and Shakhnovich, E.I. (2006). Common motifs and topological effects in the protein folding transition state. J Mol Biol *359*, 1075-1085.

Islam, S.A., Karplus, M., and Weaver, D.L. (2002). Application of the diffusion-collision model to the folding of three-helix bundle proteins. J Mol Biol *318*, 199-215.

Itzhaki, L.S., Otzen, D.E., and Fersht, A.R. (1995). The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: evidence for a nucleation-condensation mechanism for protein folding. J Mol Biol *254*, 260-288.

Jacob, J., Krantz, B., Dothager, R.S., Thiyagarajan, P., and Sosnick, T.R. (2004). Early collapse is not an obligate step in protein folding. J Mol Biol *338*, 369-382.

Jeng, M.F., Englander, S.W., Elove, G.A., Wand, A.J., and Roder, H. (1990). Structural description of acid-denatured cytochrome c by hydrogen exchange and 2D NMR. Biochemistry *29*, 10433-10437.

Jennings, P.A., and Wright, P.E. (1993). Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. Science *262*, 892-896.

Juraszek, J., and Bolhuis, P.G. (2006). Sampling the multiple folding mechanisms of Trp-cage in explicit solvent. Proc Natl Acad Sci U S A *103*, 15859-15864.

Karplus, M., and Weaver, D.L. (1976). Protein-folding dynamics. Nature *260*, 404-406.

Karplus, M., and Weaver, D.L. (1994). Protein folding dynamics: the diffusion-collision model and experimental data. Protein Sci *3*, 650-668.

Khorasanizadeh, S., Peters, I.D., and Roder, H. (1996). Evidence for a three-state model of protein folding from kinetic analysis of ubiquitin variants with altered core residues. Nat Struct Biol *3*, 193-205.

Kiefhaber, T. (1995). Kinetic traps in lysozyme folding. Proc Natl Acad Sci U S A *92*, 9029-9033.

Kim, P.S., and Baldwin, R.L. (1982). Influence of charge on the rate of amide proton exchange. Biochemistry *21*, 1-5.

Kim, P.S., and Baldwin, R.L. (1990). Intermediates in the folding reactions of small proteins. Annu Rev Biochem *59*, 631-660.

Kimura, T., Maeda, A., Nishiguchi, S., Ishimori, K., Morishima, I., Konno, T., Goto, Y., and Takahashi, S. (2008). Dehydration of main-chain amides in the final folding step of single-chain monellin revealed by time-resolved infrared spectroscopy. Proc Natl Acad Sci U S A *105*, 13391-13396.

Kimura, T., Uzawa, T., Ishimori, K., Morishima, I., Takahashi, S., Konno, T., Akiyama, S., and Fujisawa, T. (2005). Specific collapse followed by slow hydrogen-bond formation of beta-sheet in the folding of single-chain monellin. Proc Natl Acad Sci U S A *102*, 2748-2753.

Lam, A.R., Borreguero, J.M., Ding, F., Dokholyan, N.V., Buldyrev, S.V., Stanley, H.E., and Shakhnovich, E. (2007). Parallel folding pathways in the SH3 domain protein. J Mol Biol *373*, 1348-1360.

Lazaridis, T., and Karplus, M. (1997). "New view" of protein folding reconciled with the old through multiple unfolding simulations. Science *278*, 1928-1931.

Lee, I.H., Kim, S.Y., and Lee, J. (2012). Folding models of mini-protein FSD-1. J Phys Chem B *116*, 6916-6922.

Leopold, P.E., Montal, M., and Onuchic, J.N. (1992). Protein Folding Funnels - a Kinetic Approach to the Sequence Structure Relationship. P Natl Acad Sci USA *89*, 8721-8725.

Levinthal, C. (1968). Are There Pathways for Protein Folding. J Chim Phys Pcb *65*, 44-45.

Lewis, P.N., Momany, F.A., and Scheraga, H.A. (1971). Folding of polypeptide chains in proteins: a proposed mechanism for folding. Proc Natl Acad Sci U S A *68*, 2293-2297.

Lindberg, M., Tangrot, J., and Oliveberg, M. (2002). Complete change of the protein folding transition state upon circular permutation. Nat Struct Biol *9*, 818-822.

Lustig, B., and Fink, A.L. (1992). Secondary structure formation precedes tertiary structure in the refolding of ribonuclease A. Biochim Biophys Acta *1121*, 229-233.

Magg, C., and Schmid, F.X. (2004). Rapid collapse precedes the fast two-state folding of the cold shock protein. J Mol Biol *335*, 1309-1323.

Mann, C.J., and Matthews, C.R. (1993). Structure and stability of an early folding intermediate of Escherichia coli trp aporepressor measured by far-UV stopped-flow circular dichroism and 8-anilino-1-naphthalene sulfonate binding. Biochemistry *32*, 5282-5290.

Matouschek, A., Kellis, J.T., Jr., Serrano, L., and Fersht, A.R. (1989). Mapping the transition state and pathway of protein folding by protein engineering. Nature *340*, 122-126.

Matthews, C.R. (1993). Pathways of protein folding. Annu Rev Biochem *62*, 653-683.

McCallister, E.L., Alm, E., and Baker, D. (2000). Critical role of beta-hairpin formation in protein G folding. Nat Struct Biol *7*, 669-673.

Mok, K.H., Kuhn, L.T., Goez, M., Day, I.J., Lin, J.C., Andersen, N.H., and Hore, P.J. (2007). A pre-existing hydrophobic collapse in the unfolded state of an ultrafast folding protein. Nature *447*, 106-109.

Munoz, V., and Serrano, L. (1996). Local versus nonlocal interactions in protein folding and stability—an experimentalist's point of view. Fold Des *1*, R71-77.

Myers, J.K., and Oas, T.G. (1999). Contribution of a buried hydrogen bond to lambda repressor folding kinetics. Biochemistry *38*, 6761-6768.

Nolting, B., and Agard, D.A. (2008). How general is the nucleation-condensation mechanism? Proteins *73*, 754-764.

Onuchic, J.N., Luthey-Schulten, Z., and Wolynes, P.G. (1997). Theory of protein folding: the energy landscape perspective. Annu Rev Phys Chem *48*, 545-600.

Onuchic, J.N., Nymeyer, H., Garcia, A.E., Chahine, J., and Socci, N.D. (2000). The energy landscape theory of protein folding: insights into folding mechanisms and scenarios. Adv Protein Chem *53*, 87-152.

Pande, V.S., Grosberg, A., Tanaka, T., and Rokhsar, D.S. (1998). Pathways for protein folding: is a new view needed? Curr Opin Struct Biol *8*, 68-79.

Park, S.H., O'Neil, K.T., and Roder, H. (1997). An early intermediate in the folding reaction of the B1 domain of protein G contains a native-like core. Biochemistry *36*, 14277-14283.

Plotkin, S.S., and Onuchic, J.N. (2002a). Understanding protein folding with energy landscape theory. Part I: Basic concepts. Q Rev Biophys *35*, 111-167.

Plotkin, S.S., and Onuchic, J.N. (2002b). Understanding protein folding with energy landscape theory. Part II: Quantitative aspects. Q Rev Biophys *35*, 205-286.

Pradeep, L., and Udgaonkar, J.B. (2002). Differential salt-induced stabilization of structure in the initial folding intermediate ensemble of barstar. J Mol Biol *324*, 331-347.

Ptitsyn, O.B. (1973). Stages in the mechanism of self-organization of protein molecules. Dokl Akad Nauk SSSR *210*, 1213-1215.

Qi, P.X., Sosnick, T.R., and Englander, S.W. (1998). The burst phase in ribonuclease A folding and solvent dependence of the unfolded state. Nat Struct Biol *5*, 882-884.

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

Raschke, T.M., and Marqusee, S. (1997). The kinetic folding intermediate of ribonuclease H resembles the acid molten globule and partially unfolded molecules detected under native conditions. Nat Struct Biol *4*, 298-304.

Sadqi, M., Lapidus, L.J., and Munoz, V. (2003). How fast is protein hydrophobic collapse? Proc Natl Acad Sci U S A *100*, 12117-12122.

Sali, A., Shakhnovich, E., and Karplus, M. (1994). How does a protein fold? Nature *369*, 248-251.

Schmid, F.X., and Baldwin, R.L. (1979). Detection of an early intermediate in the folding of ribonuclease A by protection of amide protons against exchange. J Mol Biol *135*, 199-215.

Serrano, L., Kellis, J.T., Jr., Cann, P., Matouschek, A., and Fersht, A.R. (1992). The folding of an enzyme. II. Substructure of barnase and the contribution of different interactions to protein stability. J Mol Biol *224*, 783-804.

Sinha, K.K., and Udgaonkar, J.B. (2009). Early events in protein folding. Curr Sci India *96*, 1053-1070.

Sosnick, T.R., Shtilerman, M.D., Mayne, L., and Englander, S.W. (1997). Ultrafast signals in protein folding and the polypeptide contracted state. Proc Natl Acad Sci U S A *94*, 8545-8550.

van Nuland, N.A., Chiti, F., Taddei, N., Raugei, G., Ramponi, G., and Dobson, C.M. (1998). Slow folding of muscle acylphosphatase in the absence of intermediates. J Mol Biol *283*, 883-891.

Vidugiris, G.J., Markley, J.L., and Royer, C.A. (1995). Evidence for a molten globule-like transition state in protein folding from determination of activation volumes. Biochemistry *34*, 4909-4912.

Wallin, S., Zeldovich, K.B., and Shakhnovich, E.I. (2007). The folding mechanics of a knotted protein. J Mol Biol *368*, 884-893.

Welker, E., Maki, K., Shastry, M.C., Juminaga, D., Bhat, R., Scheraga, H.A., and Roder, H. (2004). Ultrarapid mixing experiments shed new light on the characteristics of the initial conformational ensemble during the folding of ribonuclease A. Proc Natl Acad Sci U S A *101*, 17681-17686.

Wolynes, P.G., Onuchic, J.N., and Thirumalai, D. (1995). Navigating the folding routes. Science *267*, 1619-1620.

Wright, C.F., Lindorff-Larsen, K., Randles, L.G., and Clarke, J. (2003). Parallel protein-unfolding pathways revealed and mapped. Nat Struct Biol *10*, 658-662.

Zhou, R., Huang, X., Margulis, C.J., and Berne, B.J. (2004). Hydrophobic collapse in multidomain protein folding. Science *305*, 1605-1609.

Zwanzig, R., Szabo, A., and Bagchi, B. (1992). Levinthal's paradox. Proc Natl Acad Sci U S A *89*, 20-22.