Folding consensus?

Robert L. Baldwin

Two new studies address contradictions about the use of the hierarchic mechanism for the folding process in both two-state and three-state folding reactions.

Do proteins fold by a single mechanism or do separate classes of proteins fold by different mechanisms? This question has fascinated the protein folding community for a decade, particularly after the discovery of small 'two-state' proteins which fold without observable intermediates (see ref. 1 for review). Their behavior is in sharp contrast to the folding reactions of 'three-state' proteins which display transient folding intermediates. Several three-state folders clearly fold by the hierarchic mechanism in which folding begins in the polypeptide backbone (ref. 2 and reviewed in refs 3,4), forming first local structures of only marginal stability, which can then interact with more distant partners. On the other hand, early work on the mechanism of two-state folding by CI2 suggested that the process begins with a tertiary nucleation event[5,6] which, if strictly correct, would imply that its folding begins in a nonhierarchical manner. New light is thrown on these issues by two papers, one[7] on page 151 of this issue and the other[8] in the January issue of *Nature Structural Biology*. The transient formation of excess α-helix in the folding reaction of β-lactoglobulin (βLG), a predominantly β-sheet protein, is found to be essentially a side issue, not a major barrier to the main folding process[7], while the populated transient intermediate in the folding reaction of Im7 is found to be on-pathway, not off-pathway, by measuring the kinetics in conditions in which the folding reactions of the intermediate and the native state are closely coupled[8].

β-lactoglobulin folding

The folding mechanism of β-lactoglobulin has been a prime source of concern with regard to the generality of the hierarchic mechanism. When the folding kinetics of βLG are monitored by stopped-flow circular dichroism (CD), an overshoot occurs at wavelengths used to observe α-helix formation[9,10]. The results indicate that substantially more helix is formed at early times than is present in the final native structure. βLG has a 'clam' or 'sandwich' structure with nine antiparallel β-strands, one major helix and four short helices. According to the CD-monitored folding kinetics, early events include the formation of substantial non-native helix, which must be converted later either into β-strands or nonsecondary structure. This is in contrast to the usual formulation of hierarchic folding which specifies that only native-like structure is stabilized by the folding process[3,4,11].

The nature of the folding process of βLG has now been analyzed by Kuwata, Shastry *et al.*[7]. They use submillisecond mixing, with folding monitored by tryptophan fluorescence, in a continuous-flow instrument[12]. This enables them to measure the first folding event (140 μs) detected by Trp fluorescence. Formation of the native structure takes 330 s, so the folding process spans six orders of magnitude in time. The authors obtain structural information about secondary structure present at various stages of folding by stopped-flow hydrogen-deuterium (H/D) exchange, combined with 2D ^{1}H-NMR analysis after folding is complete to determine the locations of protected peptide NH protons. The relative locations of protected peptide NH protons in the amino acid sequence of a helix *versus* a β-hairpin are quite different. The exchange results allow the authors to detect a particular secondary structure in a transient intermediate when it occurs at the same location as in the native protein and to make inferences about secondary structures present transiently at other locations. Importantly, the protection factor of each protected proton can be determined, providing a direct measure of the stability of the structure that affords protection against exchange.

The authors find a core of stably protected peptide NH protons within 2 ms of refolding, located within the F, G and H strands of the native β-sheet and also in the native α-helix near the C-terminus. They also find some weakly protected NH protons close to the N-terminus, within the segment 12–21, and they suggest that these protons are protected by a non-native helix early in folding. The 12–21 polypeptide segment is known to have a good helix propensity in peptide studies[13] and a simulation of secondary structure early in the folding of βLG, before tertiary structure is formed[4], predicts that the last part of the 12–21 segment is helical. The authors cannot find as many protected NH protons as are predicted from CD-monitored folding kinetics, and they suggest that some segments observed as helical by CD have too low stability to be detected by H/D exchange. Although the interplay during folding between the FGH core and the putative non-native helix is complex, the exchange results make a clear distinction between weakly protected non-native structure and stably protected core structure The authors interpret their results to mean that the folding process of βLG is dominated by the hierarchic mechanism, but that weakly stable, non-native helices form transiently in polypeptide chain segments which have not yet assumed their native structure in the hierarchic folding process.

Im7 folding

Some simulations of protein folding suggest that folding intermediates result from kinetic traps in which partly folded species accumulate because of barriers to further folding, while the actual folding process that forms the native structure bypasses these traps[14,15]. Other folding simulations suggest the opposite, namely that the major folding process does proceed *via* populated folding intermediates[16]. This controversy has focused interest on the problem of proving experimentally whether the populated folding intermediates seen in three-state folding reactions are on-pathway or off-pathway. Simple calculations reveal that, if the intermediate is formed within the dead time of the kinetic method (usually stopped-flow, with a dead time of 2–5 ms), it is practically impossible to decide by conventional kinetic methods alone whether the intermediate is on-pathway or off-pathway.

The situation changes if the kinetics of folding and unfolding the intermediate are measurable, and the introduction of submillisecond mixing devices has made it possible to measure the folding/unfolding kinetics of submillisecond intermediates in some protein folding reactions[12,17,18]. Even

so, the conditions for deciding from the kinetics alone whether an intermediate is on-pathway or off-pathway are quite restrictive, as Capaldi, Shastry and coworkers describe in a new submillisecond mixing study[8] of Im7, a four-helix bundle protein. Not only must the intermediate accumulate within a measurable time range, the rates of forming the intermediate and the native state must be close enough to provide kinetic coupling between the two reactions. Then, by extending the time window into the 100 μs range and by using urea to vary the relative rates of the two reactions (thereby changing the extent of accumulation of the intermediate from high to low) the authors are able to decide between the on-pathway and off-pathway mechanisms by modeling. Their results clearly show that the off-pathway model cannot fit the data while the on-pathway model does fit satisfactorily over the entire range of urea concentration. Thus the older concepts of 'folding pathway' and 'productive folding intermediate' remain useful today, although there is good evidence for competing pathways being used simultaneously in several folding reactions (see for example refs 19–21), a phenomenon that emphasizes the usefulness of the 'energy landscape' view of folding (for review, see ref. 22).

Two basic facts about the probability of molten globule intermediates being on-pathway or off-pathway have been evident for some years. First, the structures of molten globule intermediates (for review, see ref. 23) resemble those of the corresponding native proteins in important respects, including native-like (although weak) tertiary side chain interactions. This conclusion means that, although kinetic molten globule intermediates might be caught in traps caused by barriers to further folding, the intermediates nevertheless result from a folding process that produces native structure. Second, some kinetic molten globule intermediates are stable at equilibrium in non-native conditions (for review, see ref. 23). Consequently, they accumulate transiently during folding in native conditions because, first and foremost, they are stable relative to the denatured forms present at the start of folding. Of course, the kinetic barrier posed by the difficulty of reaching the folding transition state is a factor governing the extent of accumulation of the intermediate.

The results in the two new papers[7,8] are a tribute to the usefulness of the microcapillary, ultrarapid mixing method[12] used in this work. By extending the time range to 100 μs, the authors are able to include all major folding events detected by Trp fluorescence in these two folding reactions. Moreover, by introducing a mixing method that is generally applicable, they should be able to tackle whatever folding problem comes along. The value of characterizing secondary structure, as well as measuring its stability, by pulsed and competition H/D exchange is also well demonstrated in the βLG study[7].

Tertiary nucleation?

Now that Kuwata, Shastry *et al.*[7] have clarified the role of the hierarchic mechanism in the folding of βLG, what can be said about tertiary nucleation *versus* the hierarchic mechanism in two-state folding reactions? New light has been shed on this issue from an unexpected source. Plaxco *et al.*[24] have examined the folding rates of various two-state folders with the aim of finding structural properties that correlate with folding rate. By employing folding rates measured at constant stability, they find a surprisingly good correlation between log folding rate and native-state topology. The variable used to express the effect of topology is the contact order, the average distance in the amino acid sequence between pairs of residues that make contact, normalized by sequence length. The correlation between log folding rate and contact order spans an impressive million-fold range of folding rates. Because the folding rate depends by definition on structure present in the folding transition state, the success of this correlation implies that the transition state structure depends on the topology of a substantial part of the native protein, and not just on a limited nucleus. In recognition of this point, Fersht has replaced his earlier 'nucleation-condensation' model[5] for the transition state in two-state folding with an 'extended nucleus' model[25].

Other recent developments implicate the hierarchic mechanism in forming the transition state structure in two-state folding. One of the two β-turns of protein L is present in the folding transition state, according to mutational evidence, and recent NMR experiments show that specifically this turn can be detected at a very early stage in folding, when the denatured protein is examined in 2 M guanidinium chloride[26]. Thus, backbone structure formed early in folding is present in the transition state of this two-state folder. A second example is provided by the effects of certain mutations on the rate of forming the 14–38 disulfide bond in reduced, unfolded BPTI[27]. The results indicate that the disulfide closure rate depends sensitively on the stability of a β-hairpin containing a cluster of nonpolar side chains present in the backbone structure of the native protein. The fractional amount of the β-hairpin is quite small in reduced BPTI, and probably it could not be detected by standard spectroscopic techniques. Results such as these are reminiscent of Anfinsen's use of antibody detection[28] to search for minute levels of 'native format' in denatured proteins. In the hierarchic mechanism of folding, initial folding events may individually give very low levels of structure but nevertheless be mutually reinforcing when two folding events are coupled, so that the first observed folding event is likely to depend on the sensitivity of the assay used to test for it.

Folding simulations

Progress in the simulation of folding has resulted from the use of the hierarchic mechanism to predict transition state structures and folding rates. At least four different research groups have recently used models based on the hierarchic mechanism for predictions of this kind, with some success (for review, see ref. 29). Objections are now fading fast to the concept that both two-state and three-state folding reactions occur by the hierarchic mechanism. Whether the hierarchic mechanism will also apply to currently mysterious folding reactions such as amyloid fiber formation, when they are elucidated, remains to be seen.

Robert L. Baldwin is in the Department of Biochemistry, Beckman Center, Stanford University School of Medicine, Stanford, California 94305-5307, USA. email: bbaldwin@cmgm.stanford.edu

1. Jackson, S.E. *Folding & Design* **3**, R81–R91 (1998).
2. Raschke, T.M, Kho, J. & Marqusee, S. *Nature Struct. Biol.* **6**, 825–831 (1999).
3. Baldwin, R.L. & Rose, G.D. *Trends Biochem. Sci.* **24**, 26–33 (1999).
4. Baldwin, R.L & Rose, G.D. *Trends Biochem. Sci.* **24**, 77–83 (1999).
5. Fersht, A.R. *Proc. Natl. Acad. Sci. USA* **92**, 10869–10873 (1995).
6. Itzhaki, L.S., Otzen, D.E. & Fersht, A.R. *J. Mol. Biol.* **254**, 260–288 (1995).
7. Kuwata, K. *et al. Nature Struct. Biol.* **8**, 151–155 (2001).
8. Capaldi, A.P. *et al. Nature Struct. Biol.* **8**, 68–72 (2001).
9. Kuwajima, K., Yamaya, H. & Sugai, S. *J.Mol. Biol.* **264**, 806–822(1996).
10. Hamada, D., Segawa, S. & Goto, Y. *Nature Struct. Biol.* **3**, 868–873 (1996).
11. Ptitsyn, O.B. *Dokl. Akad. Nauk. SSSR* **210**, 1213–1215 (1973).
12. Shastry, M.C.R., Luck, S.D. & Roder, H. *Biophys. J.* **74**, 2714–2721 (1998).
13. Hamada, D., Kuroda, Y., Tanaka, T. & Goto, Y. *J. Mol. Biol.* **254**, 737–746 (1995).
14. Bryngelson, J.D., Onuchic, J.N., Socci, N.D. & Wolynes, P.G. *Proteins: Struct. Funct. Genet.* **21**, 167–195 (1995).
15. Dill, K.A. & Chan, H.S. *Nature Struct. Biol.* **4**, 10–19 (1997).
16. Pande, V.S. & Rokhsar, D.S. *Proc. Natl. Acad. Sci. USA* **96**, 1273–1278 (1999).
17. Takahashi, S. *et al. Nature Struct. Biol.* **4**, 44–50 (1997).

18. Jamin, M., Yeh, S-R., Rousseau, D.L. & Baldwin, R.L. *J. Mol. Biol.* **292**, 731–740 (1999).
19. Roder, H., Elöve, G.A & Englander, S.W. *Nature* **335**, 700–704 (1988).
20. Radford,S.E., Dobson, C.M. & Evans, P.A. *Nature* **358**, 302–307 (1992).
21. Zaidi, F.N., Nath, U. & Udgaonkar, J.B. *Nature Struct. Biol* .**4**, 1016–1024 (1997).
22. Dinner, A.R. *et al. Trends Biochem. Sci.* **25**, 331–339 (2000).
23. Kuwajima, K. & Arai, M. In *Mechanisms of protein folding, second edition.* (ed. Pain, R.H.)138–174 (Oxford University Press, Oxford, UK; 2000).
24. Plaxco, K.W., Simons, K.T & Baker, D. *J. Mol. Biol.* **277**, 985–994 (1998).
25. Fersht, A.R. *Proc. Natl. Acad. Sci. USA* **97**, 1525–1529 (2000).
26. Yi, Q., Scalley-Kim, M.L., Alm, A.J. & Baker, D. *J. Mol. Biol.* **299**, 1341–1351 (2000).
27. Zdanowski, K. & Dadlez, M. *J. Mol. Biol.* **287**, 433–445 (1999).
28. Sachs, D.H., Schechter, A.N., Eastlake, A & Anfinsen, C.B. *Proc. Natl. Acad. Sci. USA* **69**, 3790–3794 (1972).
29. Baker, D. *Nature* **405**, 39–42 (2000).

It takes two to tango

Robert E. Blankenship

Transient complexes between electron transfer proteins are essential to the function of biological energy storage systems. A new structure between ferredoxin and one of its reaction partners, ferredoxin-NADP⁺ reductase, reveals how these proteins interact with each other.

Transient noncovalent associations between electron transfer proteins are essential for the function of energy-conserving systems such as photosynthesis and respiration. A cycle of docking, electron transfer, dissociation and diffusion takes place, with the specificity and kinetic competence of the electron transfer chain controlled partly by the details of the noncovalent interactions among the various reaction partners[1]. While the three-dimensional structures of many redox proteins are known, very few structures of the complexes that they form with each other have been determined. On page 117 of this issue of *Nature Structural Biology*, Kurisu *et al.*[2] report the structure of a complex between two components of the photosynthetic electron transfer chain, ferredoxin (Fd) and ferredoxin-NADP⁺ reductase (FNR).

A large literature exists on the nature of the intermolecular interactions that determine and define transient electron transfer complexes, beginning with the pioneering work that defined the ring of positively charged lysine residues that surround the heme in many *c*-type cytochromes, orienting it and favoring its docking reaction with negatively charged reaction partners[3]. Chemical modifications and site-directed mutations of the reaction partners coupled with kinetic analysis of electron transfer rates and theoretical modeling have been instrumental in mapping out in considerable detail the molecular interactions that give rise to specificity in many electron transfer systems[1]. However, in all but a handful of cases these intermolecular complexes have not been directly observed.

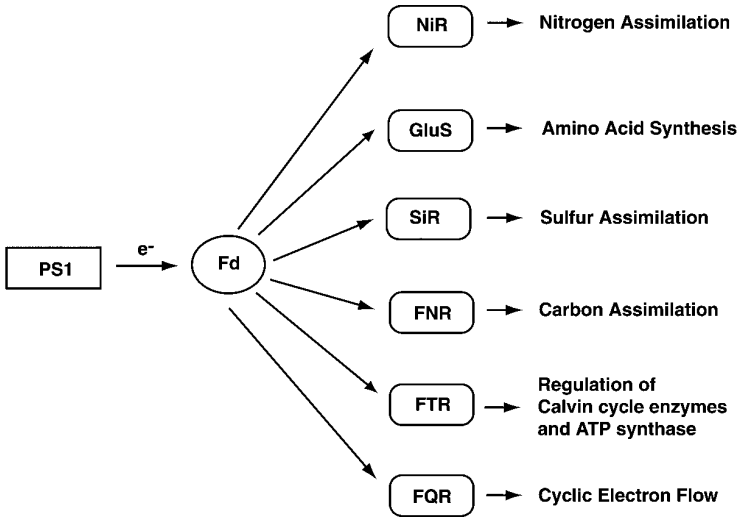

Fig. 1 Reaction network of plant-type ferredoxin (Fd) with its reaction partners, including photosystem 1 (PS1), nitrite reductase (NiR), glutamate synthase (GluS), sulfite reductase (SiR), ferredoxin-NADP⁺ reductase (FNR), ferredoxin-thioredoxin reductase (FTR), ferredoxin-plastoquinone reductase (FQR).

The first of these complexes to have its structure determined was cytochrome *c* and cytochrome *c* peroxidase[4]. Others include the ternary complex of methylamine dehydrogenase, the copper protein amicyanin and cytochrome *c*-551 in methylotrophic bacteria[5], and the cytochrome P450 complex with its flavin-containing reductase[6]. Electrostatic interactions, including salt bridges, are common in these intermolecular complexes, but they also include significant hydrogen bonding and hydrophobic components. The observed complexes usually contain fewer specific interactions than had been implicated by biochemical and kinetic studies, suggesting that the complex that crystallizes is only one of a family of kinetically competent complexes.

One of the most important of the soluble electron transfer proteins in oxygenic photosynthesis is the 2-Fe, 2-S soluble ferredoxin (Fd)[7,8]. This protein, known as plant-type ferredoxin (although an almost identical protein is found in cyanobacteria and algae) to distinguish it from the 4-Fe, 4-S bacterial ferredoxins, is a small highly acidic protein of ~11 kDa. It receives electrons from the integral membrane photosystem 1 reaction center complex and delivers them to the flavoprotein ferredoxin-NADP⁺ reductase (FNR). The first electron transfer produces a stable FAD flavin semiquinone in FNR, which is reduced again by another Fd and in turn reduces