Annu. Rev. Biochem. 1990. 59:631–60

INTERMEDIATES IN THE FOLDING REACTIONS OF SMALL PROTEINS[1]

Peter S. Kim

Whitehead Institute for Biomedical Research, Nine Cambridge Center, Cambridge, Massachusetts 02142 and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Robert L. Baldwin

Department of Biochemistry, Beckman Center, Stanford University School of Medicine, Stanford, California 94305

KEY WORDS: protein folding, disulfide bonds, hydrophobic interactions, peptide models, proline isomerization.

CONTENTS

[1]Abbreviations used: α-LA, α-lactalbumin; BPTI, bovine pancreatic trypsin inhibitor; CD, circular dichroism; cyt c, cytochrome c (horse heart unless noted differently); DHFR, dihydrofolate reductase; GuHCl, guanidinium chloride; H-bond, hydrogen bond; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; PPIase, porcine prolyl *cis-trans* isomerase (cyclophilin); RNaseA, bovine ribonuclease A; RNaseS, bovine ribonuclease S; SNase, staphylococcal nuclease; T_m, temperature at the midpoint of a thermal unfolding transition.

0066-4154/90/0701-0631$02.00

PERSPECTIVES AND SUMMARY

The basic mechanisms of information transfer from DNA to RNA (transcription) and from RNA to proteins (translation) are understood at least in outline. In contrast, understanding the transfer of information from one dimension to three dimensions (i.e. protein folding) remains as a major unsolved problem in modern molecular biology.

For small, single-domain proteins, the protein folding problem can be reduced to one of physical chemistry, whereby a unique conformation of a precisely defined solute (the polypeptide chain) is obtained when it is placed in a defined solvent (aqueous buffer) in the absence of other cofactors or catalysts. Theoretical attempts to model the protein folding process from first principles are, however, severely hindered because of the enormous number of competing interactions to consider, and because the difference in free energy between the folded and unfolded states is a tiny fraction of the energies favoring and opposing folding: the calculations must necessarily be very accurate.

Following one of the basic rules of biochemistry is proving to be an effective way to attack the protein folding problem from an experimental viewpoint. The rule is to identify intermediates in the process and then to characterize these intermediates. In 1982, when we last reviewed this subject (1), we concluded that folding intermediates certainly exist and that methods would probably be found soon for characterizing their structures. That optimistic forecast turned out to be well founded, and much attention is now aimed at determining the structures of folding intermediates, recognizing that any account of protein folding will be obviously incomplete until it includes a structural description. The focus is still on small, single-domain proteins: protein folding constitutes a complex problem and it will probably be solved by studying the simplest case first. Obtaining additional structural information about folding intermediates remains as a major task for the future, but the available data indicate that these structures are remarkably nativelike, suggesting that a large part of the puzzle is to understand protein structural hierarchies.

The problem of how to describe a folding pathway in formal kinetic terms has not yet been resolved. The simplest description is preferred, and thus the simple sequential model (1) is likely to be used until compelling evidence is presented that a more complex model is needed. The meaning of a sequential pathway for folding is analogous to a sequence of intermediates in a metabolic pathway: to go from I_1 to I_3, it is necessary to proceed though I_2. Reasons for doubting the validity of the simple sequential model are based on physical reasoning that alternative pathways of folding must be available. In some cases, there is evidence to suggest that an alternative pathway for folding exists, analogous to a salvage pathway in metabolism, but even in these cases the available data can be explained using no more than two sequential pathways. The evidence is good that pathways representing kinetically preferred routes for folding can be described: other routes may exist, but they are not depicted because they are less probable.

The transitions between different intermediates probably involve significant activation barriers and, therefore, distinct intermediates are observed rather than a continuum. These transitions appear to involve structure-determining steps, such as the formation of a disulfide bond, that "lock" a region of the protein into a particular folded state. At present, these "adhesion steps" can only be guessed at, from inspection of the X-ray structures of proteins. Identifying adhesion steps as they occur in folding intermediates is a major task for the immediate future.

The ability to create readily single-amino-acid substitutions using recombinant methods, together with improved methods in NMR and X-ray crystallography, has led to a dramatic increase in the amount of data relating amino acid sequence to three-dimensional structure. The general picture to emerge is that proteins are remarkably tolerant to amino acid substitutions: mutations are often accommodated by minor adjustments of protein structure without a substantial decrease in overall stability [reviewed recently by Alber (2)]. A vivid demonstration of the plasticity of proteins comes from Lim & Sauer (3), who find that there are a remarkable number of non-native combinations of hydrophobic core residues that allow lambda repressor to fold properly. Such mutagenesis studies emphasize the degenerate nature of the "code" for folding, a point made earlier by analysis of the sequences that give the globin fold (4). Nevertheless, recent success in the de novo design of proteins indicates that the folding problem does not necessarily have to be complex.

SCOPE OF THIS REVIEW

We focus here on recent developments in understanding the folding of small, single-domain proteins. The central topics are structures of folding inter-

mediates and pathways of folding. There are numerous alternative approaches, not reviewed here, to the larger problem of understanding the relationship between amino acid sequence and three-dimensional structure. In particular, significant improvements in homology alignment procedures have occurred, and work on the energetics of folding and the design of new structures is proceeding rapidly.

The protein folding problem has just been reviewed for both the initiated and uninitiated reader by King (5). A comprehensive review of studies on the folding process with emphasis on oligomeric proteins has been given by Jaenicke (6). Work from Scheraga's group on the formation of local structures in protein folding has been reviewed recently (7). Ptitsyn's review (8) of the folding process emphasizes the molten globule model for folding intermediates. There are books by Lapanje (9) on protein denaturation and by Ghélis & Yon (10) on protein folding, and the book on proteins by Creighton (11) provides an excellent summary of background material. Three long reviews of calorimetric investigations of the folding process by Privalov (12, 13) and Privalov & Gill (14) could easily make another book. The Richardsons (15) and also DeGrado (16) have recently reviewed design of novel proteins. The measurement and thermodynamic analysis of protein stability have been reviewed by Schellman (17). Goldenberg (18) has reviewed studies based on the genetic approach to the folding problem. Chothia (19) has summarized work on the principles of folding as they are revealed by protein crystal structures. Theoretical studies of folding have been reviewed by Gō (20). The relationship between kinetic folding intermediates and the compact "molten globule" form is the subject of a recent review by Kuwajima (21). The initiation of protein folding is covered in two current reviews (22, 23). Principles and problems of protein folding have been reviewed by Creighton (24). A 1980 book of collected papers on protein folding, edited by Jaenicke (25), remains a valuable reference.

The role of heat-shock proteins and chaperonins in protein export and in vivo folding have been reviewed by Randall et al (26) and Rothman (27). Freedman (28) has reviewed protein disulfide isomerase, and evidence that this enzyme functions in vivo to catalyze disulfide bond formation in the endoplasmic reticulum. Inclusion body formation, a common problem with recombinant proteins, has been reviewed recently (29, 30). Taylor (31) has reviewed different methods of structure prediction and analysis of homologous sequences.

EXISTENCE OF INTERMEDIATES

In 1982, those scientists who were studying folding intermediates were convinced of their reality, but the wider community was skeptical. The

two-state model of the folding process (U⇌N, U=unfolded, N=native, without observable intermediates) was deeply engrained and implied that folding intermediates were difficult, perhaps impossible, to detect. The following developments were influential in convincing others of the reality of folding intermediates.

First, compact kinetic intermediates can be seen directly on gels (32) when the folding transition is examined by urea-gradient electrophoresis near 0°C, conditions that allow the kinetics of folding to be analyzed for some proteins.

Second, probes of secondary structure easily detect structural intermediates in kinetic folding experiments, especially when folding is studied in strongly native conditions (e.g. low temperatures). This was observed first using amide hydrogen exchange and ^3H labeling (33, 34). Later, fast CD measurements revealed the presence of early folding intermediates in kinetic studies (35, 36).

Third, an equilibrium form of some proteins has been found that is obviously different from both the native and unfolded forms and may be an intermediate in folding (8, 21). A molten globule model has been proposed (8, 37) to describe this collapsed form (see next section, on MODELS). Originally the collapsed form could be found in only a few proteins, but recently either this form or some related form has been found in other proteins (38, 39). The rapidly formed intermediates found by fast CD measurements (36) resemble the collapsed form (8, 21), suggesting that it may be on the pathway of folding.

Fourth, fragments of small proteins provide models for folding intermediates in some cases (40–43), by forming weak but folded structures, especially at low temperatures (0°C). The structures in these fragments are nativelike, and the results suggest that the entire polypeptide chain is not needed for folding to occur.

Fifth, a method for analyzing the structures of kinetic folding intermediates has been developed that combines stopped-flow pulse labeling with subsequent 2D-NMR analysis after folding is complete (44, 45), and the first results indicate that nativelike secondary structures are present in early folding intermediates of RNaseA (44) and cyt c (45). The method employs exchange with solvent (H_2O or D_2O) to label and trap exposed peptide NH protons in kinetic folding intermediates.

Finally, kinetic folding intermediates that occur in the pathway of disulfide bond formation of BPTI have been recognized since the early 1970s (46). Because the process of disulfide reoxidation is coupled to the refolding of reduced BPTI, these disulfide intermediates are also necessarily refolding intermediates. The linkage relations between folding and disulfide bond formation have been analyzed (47, 48), thus clarifying the nature of disulfide intermediates as folding intermediates.

Consequently, in this review we consider that the existence of folding intermediates has already been documented satisfactorily, and we consider questions about their nature and the pathways of folding.

MODELS FOR FOLDING

In our previous review (1), working models played an important part in correlating observations on folding intermediates. The framework model postulated that the native secondary structure is formed before the tertiary structure is locked in place. The modular assembly model proposed that even a small protein could fold by parts, and that subdomains would be important structures in folding intermediates. In this review, models have been deliberately pushed into the wings, to await their next turn on stage. The reason is that detailed structural information on folding intermediates is now becoming available. Structural studies combined with site-directed mutagenesis should identify interactions that stabilize folding intermediates. When this information is available, it may be appropriate to label the mechanisms of protein folding by a single model that expresses the central features, or these features may be diverse, like the structures of proteins.

The term "molten globule" often occurs in current literature connected with folding intermediates. Two different meanings are associated with the term, one being a model and the other being an observed form of the protein. The model given by Ptitsyn (8) specifies that the structure of the molten globule is liquidlike and fluctuating ("molten") while still being compact and having a high content of secondary structure ("globule"). We use the term "molten globule" to refer to this model. The "molten globule" form of a protein is a compact form that is widely different from both the native and unfolded forms and can often be observed at acid or alkaline pH in moderate or high salt concentrations (21, 38). In this second meaning, "molten globule" refers to a "collapsed form," and we use the latter term here for the observed form of a protein.

NATURE OF FOLDING INTERMEDIATES

The practical problem in studying folding intermediates is that the folding reactions of most single-domain proteins are highly cooperative, so that intermediates are not well populated at equilibrium. There are three general strategies that workers in the field have used to circumvent this problem. The first is to carry out kinetic refolding experiments, in which activation barriers lead to the transient accumulation of intermediates and low temperatures can be used to stabilize intermediates. The second is to find special conditions where partially folded forms of the protein are populated at equilibrium. The

third is to identify protein fragments or peptide models that fold autonomously. In addition, consideration of residual structure in the unfolded state has become an important issue in understanding early events in protein folding.

Disulfide Intermediates

When the three disulfide bonds in BPTI are reduced, the protein unfolds even in the absence of denaturants, and therefore protein folding is thermodynamically coupled to disulfide bond formation. The reduced, unfolded protein can be refolded by placing it in an oxidizing environment, and kinetic intermediates can be trapped covalently (by blocking remaining thiol groups) and then isolated by chromatography. Creighton's work on the oxidative folding of BPTI provides the most thorough description of a folding pathway, and this work has been reviewed in detail (24, 49). Here we only highlight the major characteristics of the BPTI folding pathway, and review recent developments.

There are two striking features of the predominant folding pathway for BPTI (also referred to as the rearrangement pathway). First, although there are 15 possible ways to form the first disulfide bond, one combination accounts for 60% of the single-disulfide intermediates (50). This species contains the 30–51 disulfide (found in the native protein) and is denoted [30–51]. All subsequent intermediates in the rearrangement pathway retain the 30–51 disulfide bond, emphasizing the importance of [30–51] in the folding of BPTI. A number of experiments indicate that the distribution of intermediates at the single-disulfide step is under thermodynamic control, and that [30–51] contains a nonrandom conformation (review, 49). The second striking feature is that disulfide bond formation does not involve sequential addition of native disulfide bonds, but instead folding by rearrangements of incorrect two-disulfide species is strongly preferred kinetically (51). A structural explanation for why these rearrangements occur requires an explanation for why participation by Cys-55 in the second step of folding is conspicuously absent (review, 49).

The finding that a mutant BPTI molecule lacking Cys-14 and Cys-38 (i.e. residues involved in the two-disulfide rearrangements) refolds at a rate only slightly slower than that of the wild-type protein, called into question (52) the rearrangement pathway determined by Creighton. A more detailed kinetic analysis (53) showed, however, that the rates of forming the incorrect two-disulfide intermediates in the wild-type protein were 5000-fold faster than the rate of forming the second disulfide bond in the mutant. Thus, direct formation of the second native disulfide bond is possible, but is not kinetically significant in the folding of the wild-type protein, a conclusion originally obtained (51) using chemically modified BPTI species.

Approximately half of the molecules do not enter the rearrangement pathway in strongly oxidizing refolding conditions but become trapped as a metastable state containing two native disulfides, [5–55; 14–38]: the thiols of Cys-30 and Cys-51 are buried and therefore unreactive (54). This metastable species is stable for hours at 25°C and for months at 4°C. NMR studies (54) and recent crystallographic studies of a mutant BPTI containing Ala-30 and Ala-51 (55) indicate that the metastable species has essentially the same overall conformation as the native protein. Curiously, no obvious precursor to the metastable species is populated significantly among the one-disulfide intermediates. In a detailed kinetic analysis, Creighton & Goldenberg (48) showed that the precursor goes on to form the metastable species at a rate 50-fold greater than the corresponding rate for formation of two-disulfide species from [30–51]. These workers suggested that [5–55] was the most likely precursor to the metastable species. ([5–55] is an observed intermediate in the folding of circular BPTI; Ref. 56.) Thus, a fork in the folding pathway for BPTI occurs at the one-disulfide stage: [30–51] leads to the rearrangement pathway and formation of native protein, while another intermediate (possibly [5–55]) leads to formation of a metastable species.

Because protein folding and disulfide bond formation are thermodynamically linked for BPTI, structural characterization of the trapped intermediates should allow one to understand the pathway in structural terms, provided that the blocking groups do not interfere with the conformations of the intermediates (alternatively, mutagenesis can be used to replace the blocked cysteine residues). An NMR study of the chemically trapped species, purified by chromatography, was hindered by low solubility and the presence of substantial unfolded regions in many of the intermediates (57). Nevertheless, a survey of the trapped intermediates using 1D-NMR (57) confirmed earlier conclusions that the intermediates contained nonrandom structure (49), and suggested that structure in the intermediates became more nativelike as one progressed through the pathway.

Peptide Models

Disulfide-bonded peptide models of the BPTI folding intermediates have been used recently to investigate Creighton's rearrangement pathway and formation of the metastable species. The primary advantage of these peptide models is that, by removing parts of the protein, solubility is increased substantially so that 2D-NMR can be used for structural characterization (42).

The first peptide model of a BPTI folding intermediate was for [30–51], the crucial one-disulfide intermediate in the rearrangement pathway. The model was designed by assuming that the structure in the intermediate was nativelike and localized near the disulfide bond, and consists of two short peptides connected to each other via a disulfide bond corresponding to the 30–51

disulfide in BPTI (42). This peptide model is very soluble in aqueous solution (15 mM) and contains a large amount of folded structure. A comparison of the 1D-NMR aromatic spectrum for [30–51] with that for the peptide model (42) indicates that the essential features of the structure in authentic [30–51] are captured in the peptide model.

The striking feature to emerge from 2D-NMR studies of this peptide model is that the structure is very nativelike: the C-terminal α-helix, central β-sheet, and hydrophobic core between them are folded in a manner almost indistinguishable from that found for the corresponding regions in the native protein (42; T. G. Oas, P. S. Kim, unpublished results). These results show that nativelike 2° and 3° structure can form early in folding, and provide a structural explanation for why [30–51] predominates at the single-disulfide stage of folding. Since all subsequent intermediates in the rearrangement pathway for BPTI retain the 30–51 disulfide, the nativelike structure that forms early is likely to persist and influence later steps in folding. In the peptide model, Cys-55 was replaced with an Ala residue, and NOE measurements indicate that Ala-55 (located in the α-helix) is involved in tertiary interactions with the β-sheet region (42), providing a possible explanation for why Cys-55 does not participate in the second step of the rearrangement pathway, and why the rate-limiting step in the pathway involves formation of the 5–55 disulfide bond (T. G. Oas, P. S. Kim, unpublished).

A second disulfide-bonded peptide model (43) was designed to test the hypothesis (48) that [5–55] was the precursor to the metastable species. The same assumptions were used in the design of this peptide model, and the two regions were connected using a disulfide corresponding to the 5–55 disulfide; Cys-30 and Cys-51 were replaced by Ala residues (43). This peptide model also shows substantial stability in aqueous solution, but only if regions corresponding to the central antiparallel β-sheet of the protein are included. 1D-NMR spectra suggest that the structure is again nativelike. Moreover, the aromatic ring of Phe-45 shows hindered rotation on the NMR time scale (43), analogous to that observed with intact BPTI (59). The high stability of the peptide model strongly supports the suggestion that [5–55] is the precursor to the metastable species, and the apparent nativelike folding of the peptide model suggests a structural explanation for why the second disulfide bond forms so quickly from this precursor (43).

Earlier peptide models also pointed to the importance of nativelike structure in protein folding. The most detailed studies on isolated α-helix formation by protein fragments have focused on the N-terminal α-helix of RNase A: C-peptide corresponds to residues 1–13 and S-peptide to residues 1–20. Both peptides show substantial α-helix formation, as monomers, in aqueous solution at low temperatures (60–62). Recently, a 12-residue peptide corresponding to the C-terminal α-helix of BPTI has also been found to show partial

helix formation in aqueous solution (63), and a peptide from myohemerythrin exists as a collection of rapidly interconverting turnlike structures termed the nascent helix (64).

In native RNase A, the helix is localized to residues 3–13. NMR (65, 66) and amino acid substitution (40) studies suggest that the C-peptide helix stops at the N-terminal end because a salt bridge is formed between the side chains of Glu-2^- and Arg-10^+. In the crystal structure of the native protein, this same salt bridge is observed (67), and model building (40) shows that formation of the Glu-2^-•••Arg-10^+ salt bridge prevents the backbone of residue 2 from acquiring a helical conformation. Thus, a specific side chain–side chain interaction prevents helix propagation in both the isolated peptide and the intact protein.

At the C-terminal end of the helix, a tertiary hydrogen bond between the amide proton of residue 14 (i.e. the next H-bond donor if the helix were to continue) and the carbonyl of residue 47 is observed in the crystal structure of RNase A (67). This tertiary H-bond is obviously not present in the isolated S-peptide. Nevertheless, the helix formed by the isolated S-peptide stops at or near residue 13 (62, 68). Thus, the S-peptide sequence contains a helix stop signal that functions in the absence of tertiary interactions. Nelson & Kallenbach (69) have shown that the helix stop signal in S-peptide functions in helix-promoting solvents such as trifluoroethanol/water mixtures, and Wemmer and coworkers (70) find that this helix stop signal functions even when the helix is nucleated in a hybrid peptide by using disulfide bonds.

A helix stop signal involving a proline residue at the end of a helix in T4 lysozyme has been identified by Alber, Matthews, and coworkers (71): replacement of this proline by any one of a number of residues allows the helix in the intact protein to propagate for almost one complete helical turn, although the stability of the protein remains approximately the same.

An interesting use of protein fragments to study folding has been described recently by Frieden and coworkers (72). A 53-residue fragment of DHFR, corresponding to three β-strands at the C-terminus of the protein, is capable of inhibiting folding of the intact protein. The fragment does not have a significant effect on the activity of the folded enzyme. Fontana and coworkers (73) have identified a 62-residue fragment of thermolysin, corresponding to three α-helices, that forms a stable (T_m > 60°C) folded structure in aqueous solution, even as a monomer (74).

There is one case of a peptide model that shows folded structure in aqueous solution different from that found in the native protein. Wright, Lerner, and coworkers find that a nine-residue peptide from the hemagglutinin of influenza virus forms a β-turn structure in aqueous solution (75). The crystal structure of the intact protein, however, shows that this region of the protein exists in an extended conformation (76).

Pulse Labeling

The rate of exchange between solvent protons and the backbone amide protons of proteins can be used as sensitive probes for folded structure, because H-bonding and/or protection from solvent lead to significant retardation (often $> 10^4$-fold) of exchange rates. Protection of NH protons from exchange in folding intermediates was originally demonstrated by [^{3}H]-H_2O methods; the results showed that substantial folding occurred before complete tertiary structure was obtained in kinetic refolding experiments (33, 34). Recently, the protection from exchange has been measured by using quench-flow H_2O/D_2O labeling methods, followed by subsequent 2D-NMR analysis (deuterium is not detected in proton NMR measurements) to give information about exchange at individual sites in a protein. The final analysis by NMR is performed after folding is complete so that the labeled protons are stable, and NMR assignments are needed only for the native protein. The study of folding intermediates by NMR is discussed by Roder (78).

For both RNaseA (44, 77) and cyt c (45), many amide protons are protected from exchange before folding is complete. In the case of RNaseA, the pattern of protection from exchange suggests that most or all of the β-sheet (44) and two of the three α-helices (77) are formed early in folding. For cyt c, amide protons in the N- and C-terminal helices show similar protection from exchange at early stages of folding (45), and pairing between the N- and C-terminal helices appears to be an early event in folding.

Collapsed Forms (Molten Globules)

Several years ago it was observed by Kuwajima, Sugai, and coworkers (21, 79, 80) that α-LA forms an equilibrium intermediate at low pH (originally called the A state). The CD spectrum of the A state was essentially identical to that of the native protein in the far UV region, but the aromatic region of the CD spectrum was similar to that of the completely unfolded protein. Thus, the A state appeared to be an equilibrium intermediate containing nativelike secondary structure but lacking tertiary structure (80).

The surprise came when it was found that the A state of α-LA was compact as measured by intrinsic viscosity (81) or dynamic light scattering measurements (82). These measurements indicated that the A state was much more compact than the unfolded protein (e.g. in GuHCl) but only slightly less compact than the native protein. This led to Ptitsyn's interpretation (8) of the A state as a "molten globule": a compact folding intermediate with a high content of secondary structure and fluctuating tertiary structure. The name "molten globule" was given by Ohgushi & Wada (37).

The current literature on "molten globules" is rather confusing. One source of confusion is that the same term is used for Ptitsyn's model and for the compact intermediate form that is observed experimentally. Some evidence

indicates that Ptitsyn's model may not accurately describe the experimentally observed state. In particular, the rapidly fluctuating liquidlike structure postulated for the molten globule is incompatible with a structure determined by fixed tertiary interactions. As mentioned earlier (see section on MODELS) we use the term "molten globule" for Ptitsyn's model, and the term "collapsed form" for the experimentally observed intermediate. In addition, it is not yet clear if there is a single collapsed form or a series of related forms (21).

A collapsed form has been detected in other proteins (38, 39), and it now seems likely that it occurs commonly but in restricted conditions [e.g. in acid or alkaline pH, at moderate or high salt conditions (38)]. Collapsed forms have a strong tendency to aggregate (21, 38), making it difficult to characterize them, and emphasizing the importance of determining the aggregation state in any study of collapsed forms. The use of acidic or alkaline conditions to obtain collapsed forms can lead to chemical damage of the protein (38), and it is important to demonstrate that this has not occurred.

In kinetic refolding experiments, fast CD measurements indicate that rapidly formed intermediates have at least one property resembling the collapsed form (36). This suggests that the collapsed form may be on the pathway of folding. The equilibrium between the collapsed form and the unfolded state is fast ($\tau < 1$ msec), whereas the equilibrium between the collapsed form and the native protein is slow ($\tau > 1$ sec) (21). The lack of a large activation barrier between the collapsed form and the unfolded state suggests that the collapsed form may correspond to the structure of the unfolded protein in refolding conditions.

Calorimetric studies (83, 84) of α-LA suggest that the collapsed form has nearly the same heat capacity as the unfolded state (21, 85). Since the increase in heat capacity that accompanies protein folding is thought to result from the exposure of previously buried nonpolar side chains to water (12, 86), this result leads to the surprising conclusion that nonpolar side chains in the collapsed form are solvated (the uncertainty in the measurements on α-LA, however, would allow the Cp of the collapsed form to be 20% on the way to native protein). Collapsed forms bind hydrophobic dyes strongly (21, 38), at concentrations where the native or unfolded forms of the protein do not bind, suggesting that collapsed forms contain accessible hydrophobic surfaces. These findings, together with the observation that some collapsed forms aggregate readily, raise the possibility that they may be produced by a collapse involving hydrophobic interactions between solvated side chains.

Recent NMR studies of α-LA provide the first direct structural information about a collapsed form (87; P. A. Evans, C. M. Dobson, personal communication). The NMR data indicate that some fixed structure is in fact present in this collapsed form, in contrast to the hypothesis that structure in the molten globule is rapidly fluctuating. At least two α-helices that are

present in the native protein are also present in the collapsed form, and peptide NH protons in these helices are well protected from exchange. In contrast, the β-sheet regions are not formed stably in the collapsed form of α-LA (P. A. Evans, C. M. Dobson, personal communication). Recent NMR studies of a collapsed form of cyt c also indicate that some NH protons are protected significantly from exchange (S. W. Englander, personal communication). It is hard to imagine how amide protons could be protected substantially from exchange in the absence of specific tertiary structure. These initial results suggest that the collapse may not be uniform and that subdomain structures, with fixed tertiary elements, may be present in collapsed forms. Further NMR studies, and determination of the tertiary interactions that are present in collapsed forms, are needed to clarify these issues.

Hydrophobic Collapse

The notion of a hydrophobic collapse is based on the idea that when an unfolded polypeptide with exposed hydrophobic side chains is placed in an aqueous solution, it will collapse to a state that shields hydrophobic groups from solvent. It seems likely that a hydrophobic collapse is at least partially involved in collapsed forms, although this remains to be established firmly. Dill and coworkers (90, 91) have used lattice models to argue that a large fraction of compact states contain substantial amounts of secondary structure. Bowie & Sauer (92) give evidence that the hydrophobic nature of allowable amino acid substitutions in a protein provides substantial predictive information about secondary structure. Model compound studies suggest (see section on ENERGETICS OF FOLDING) that pairwise hydrophobic interactions can occur even when nonpolar side chains are separated by a layer of water molecules (93–96).

Haas and coworkers have used energy transfer measurements to determine distributions of distances between fluorescent groups placed on individual residues in BPTI. Whereas the distributions of distances are narrow in the native state, and are in agreement with those predicted from the crystal structure of BPTI (97), the distances measured in the reduced form of the protein are significantly shorter than predicted for a flexible chain (98), suggesting that the reduced form of BPTI is compact. Morcover, some of the distance measurements in the reduced form show a bimodal distribution of distances, indicating slow interconversion on the fluorescence time scale (98). In contrast, chemical reactivity (99), and CD (100) studies of reduced BPTI show no evidence for structure.

Studies of a series of deletion mutations in SNase have led to an interesting demonstration of a collapsed denatured state by Shortle and coworkers (101). Deletion of a few residues from the C-terminus of SNase prevents formation

of the native structure and thereby permits analysis of the denatured form at low denaturant concentrations. A nonrandom conformation is observed for these C-terminal deletion proteins. The structure is compact as determined by hydrodynamic measurements and contains significant secondary structure as measured by CD (101). The striking observation is that amino acid mutations affect the stability of the collapsed state in a manner that is distinct from the effect of the same mutations on the stability of the native protein: in some cases the rank order is reversed. These results may provide a model system for understanding the forces that produce the collapsed form. Recent work by Matthews and coworkers (102) provides suggestive evidence for a rapid hydrophobic collapse, based on an early increase in Trp fluorescence in kinetic refolding experiments with DHFR.

Local Unfolding Reactions

Intermediates in unfolding reactions are difficult to populate even in kinetic experiments, most likely because the intermediates are not stable in unfolding conditions. Transient fluctuations in the structure of a folded protein can be detected using amide proton exchange measurements (103, 104) or by fragment complementation studies (105), and it is likely that in some cases these fluctuations correspond to local unfolding reactions.

A more direct demonstration of a local unfolding reaction comes from an NMR study of the pH-induced unfolding of RNaseS (106). The rate of interconversion between the folded and unfolded states is fast on the NMR time scale when the amide protons in the S-peptide moiety are used as probes (the chemical shift of these protons changes in a continuous manner as the pH is decreased). In contrast, the His side chain protons in the S-peptide are in slow exchange with the unfolded state. In the native protein, the His side chain is involved in tertiary contacts with the S-protein. Thus, in this local unfolding reaction, tertiary interactions appear to be maintained in the absence of secondary structure (106). In addition, kinetic experiments indicate that tertiary interactions between the S-peptide and S-protein form before the S-peptide α-helix is stablized (35, 107). These results raise the possibility that specific tertiary interactions may dictate the location of some secondary structure elements.

ADHESION STEPS

Folding decisions are context dependent, as demonstrated by examples of identical pentapeptide sequences that fold into either an α-helix or a β-sheet, depending on the protein (108). This context dependence is certainly related to the cooperativity of folding. One wants to understand how structure-determining decisions are made in the folding process.

The classic picture of a structure-determining step is a helix pairing event, where the decision to fold into a helix, and remain folded as such, is made when two helices dock properly with each other. The structure-determining step involves both the interdigitation of side chains to form a proper tertiary structure, and the removal of water from the newly formed interface. Only a few different relative orientations are used in the packing of α-helices and β-sheets (review, 19), and side chain rotamer distributions are much more limited than originally thought (109), suggesting that rules based on selected stereochemistry are involved in structure-determining steps. Provided they are associated with significant activation barriers, structure-determining steps are key components of rate-determining steps in folding. We will refer to structure-determining events that "lock" (reversibly) a region of the protein into a particular folded state as *adhesion steps*.

A clear example of an adhesion step is disulfide bond formation. Studies of the oxidative refolding of BPTI have identified general features of adhesion steps that are likely to be found in the noncovalent counterparts as well. Because it involves a covalent change, it is possible to measure accurately both the rates and equilibrium constants for disulfide bond formation that accompanies refolding, as shown notably for BPTI (48, 51). Initial disulfide bond formation in the refolding of BPTI is approximately random: the distribution of single disulfide intermediates does not reflect differences in the rate of disulfide bond formation, but rather the differential thermodynamic stabilities of the possible species (49). Thus, if a particular disulfide allows a folded structure to be formed stably, then the disulfide bond is retained preferentially. Protein folding and adhesion steps (in this case, disulfide bond formation) are thermodynamically linked processes. In BPTI, this linkage involves nativelike secondary and tertiary structures (42, 43), suggesting that a hierarchical progression of adhesion steps exists that is based on the native structure.

Although adhesion steps are reversible, they can act as kinetic traps during folding. For example, the metastable species formed in kinetic refolding experiments with BPTI (48, 54) contains two native disulfide bonds, but formation of the third native disulfide is hindered drastically because the two remaining thiols are completely buried [see also (110)]. Thus, two prior adhesion steps (acquired in a particular order) act as a kinetic trap by preventing a final adhesion step from occurring.

Adhesion steps are crucial determinants of protein folding and more specifically, of cooperativity. Noncovalent adhesion steps are just beginning to be studied by analysis of solvent viscosity effects on the kinetics of protein folding [111; see also (112)], and by analysis of helix pairing events (113–115). A major challenge ahead appears to be obtaining a detailed understanding of adhesion steps.

KINETIC PATHWAYS

Thermodynamic Control of Folding

The evidence is good that the final, three-dimensional structure of a protein is under thermodynamic, not kinetic, control. This principle, which is central to the problem of understanding folding pathways, is so surprising that many workers have reservations about its validity, and some workers still propose kinetic mechanisms for the determination of structure. Nevertheless, the principle has not been seriously disputed since it was put forward by Anfinsen (116, 117). Since refolding occurs efficiently in vitro with small, single-domain proteins, the problem of kinetic versus thermodynamic control of folding can be studied with purified proteins. If a true equilibrium is reached between unfolding and refolding, inside the folding transition zone, then the folded structure must be the thermodynamically most stable one that is accessible.

A standard test for equilibrium is to determine if the same extent of reaction is reached starting from the products as from the reactants. In the case of protein folding, the test is whether the same transition curve is found in refolding and unfolding experiments. This test is satisfied routinely in studies of small proteins. If a metastable folded structure is formed that is not in equilibrium with the unfolded form, then a hysteresis loop should be found when unfolding and refolding experiments are compared [see, e.g. cation-induced DNA condensation (118)]. Consequently, a very sensitive test for determining if true equilibrium is reached is to compare the apparent rate constants in unfolding and refolding experiments, in the same final conditions (119). They should be the same if there is no hysteresis. This test is satisfied in cases where it has been studied (120, 121). For oligomeric or multidomain proteins, it is often a challenging problem to determine whether equilibrium can be reached. In the cases studied thus far, finding in vitro conditions in which refolding does occur means that unfolding/refolding equilibrium can be reached in suitably chosen conditions.

The objection is sometimes raised that another folded structure might be more stable, but there is no pathway available for forming that structure. Forbidden structures of this kind are only a hypothetical possibility until they can be demonstrated experimentally. A more interesting problem is presented by subtilisin, which attains equilibrium between folded and unfolded forms only in the presence of its "pro" sequence, a 77-residue N-terminal segment present in prosubtilisin. The unfolded proenzyme can be expressed in *Escherichia coli,* and it folds correctly after dilution from 6 M GuHCl (122). Once the proenzyme has folded, the pro sequence is removed by self-cleavage, thus forming mature subtilisin which does not, however, refold in the same conditions used to refold prosubtilisin, or in any other conditions yet studied. When the pro sequence is added back in *trans,* a low yield of mature

subtilisin can be obtained by refolding (123; see also 123a). Consequently, mature subtilisin apparently has only kinetic stability: it has no pathway available for refolding if the pro sequence is not present. An implication of this conclusion is that only localized unfolding occurs in physiological conditions (i.e. the two-state equilibrium measured at high temperature or in the presence of a denaturant cannot be extrapolated to strongly native conditions).

Formal Description of Pathways

The finding that distinct intermediates can be well populated in kinetic refolding experiments raises the question of how properly to describe the protein folding process. The sequential description of folding pathways (1) is preferred since it is the simplest kinetic expression. A clear demonstration of a unique, sequential folding pathway has been given by Nall (124) for the folding of yeast iso-2 cyt c at alkaline pH, where an inactive but folded form of the protein is produced. In kinetic experiments at alkaline pH, a nativelike, active species appears as a transient intermediate, suggesting that the alkaline form is produced in a sequential manner that follows folding to the native form (124).

A similar example exists in the slow-folding reaction of RNase A (see section on *Proline Isomerization*). Here the evidence is good that formation of a nativelike, enzymatically active intermediate, I_N, precedes in a sequential manner the proline isomerization step that generates the native protein (125, 126). For the oxidative refolding of BPTI, it is possible to trap and quantitate the population of different intermediates in kinetic refolding and unfolding experiments (49). A self-consistent set of rate constants for the forward and reverse reactions at each step is obtained using a sequential description, containing a single major fork, of the folding process (48).

Thus, where it has been investigated carefully, the results with several proteins are consistent with a sequential pathway for folding. Although branch points and/or alternative pathways may exist, they do not appear to be abundantly populated. Most workers in the field are making the working assumption that a pathway describing the kinetically preferred routes for folding can be obtained and that other routes, although feasible, are less probable. Pulse-labeling experiments are potentially able to determine if alternative folding pathways are used simultaneously, and to detect abortive intermediates during folding (77), but these issues are not yet resolved.

Harrison & Durbin (127) go beyond the suggestion of multiple pathways and argue that protein folding resembles assembly of a jigsaw puzzle, where there is no defined pathway and any observed intermediate is not unique. There is no compelling evidence that a model of this complexity is needed, and indeed, data of the sort described above argue against the jigsaw model for folding.

Thermodynamic determination of the final folded structure (see preceding

section) may imply that the folding pathway is also thermodynamically determined: this would be true if the pathway represents the route between the most stable intermediates. At early stages of folding, intermediates on different pathways equilibrate with each other, so that the best-populated pathway is the one with the most stable intermediates. Such a mechanism would explain why protein folding is different from the assembly of a jigsaw puzzle.

Transition State for Folding

The concept of a transition state is tied to the concept of a sequential pathway with well-defined intermediates. In the jigsaw model of folding (127), there is no unique transition state. Likewise, transition-state theory is not applicable to conformational transitions of flexible polymers, according to Kramers (128), who has provided an alternative theory.

By definition, the transition state cannot be populated in any ordinary chemical reaction, and the structure of the transition state cannot be determined by direct methods. Instead, inferences about its structure are made according to the changes observed in the rate constant of a reaction when certain variables are changed. For protein folding, these variables are either mutations or environmental parameters such as pH, temperature, and the concentration of a denaturant or a specific ligand.

There are two special concerns in studying the transition state in protein folding. The first concern is that the principle of microscopic reversibility is only necessarily true if the conditions used to measure refolding and unfolding are the same, so that an unfolding pathway measured in unfolding conditions may be different from a refolding pathway measured in refolding conditions. For example, the conversion of I_N to native RNaseA is the final step of folding in strongly native folding conditions when proline isomerization is involved. I_N is not detected, however, in unfolding experiments, and essentially complete unfolding of RNase A can occur on a time scale that is fast compared to that of proline isomerization. The second concern is that folding intermediates usually are suppressed in conditions where both folding and unfolding can be measured (i.e. within the unfolding transition zone). Folding intermediates are often substantially destabilized in these conditions (moderate temperature or substantial denaturant concentration), and this entails the risk of changing the folding pathway.

Instead of suppressing folding intermediates, it is preferable to characterize them, especially intermediates present on either side of the rate-limiting step. For example, the effects of mutations on different steps in the oxidative refolding pathway of BPTI are being analyzed by Goldenberg and coworkers (129). The transition state for the folding of α-LA has been studied in the presence of a rapidly formed folding intermediate by assuming that the intermediate is in rapid equilibrium with the unfolded protein (130).

Studies of the transition state provide a reasonably consistent picture for the different proteins that have been studied: BPTI (131), hen lysozyme (132, 133), α-LA (130), a T4 lysozyme mutant (134), and barnase (135). The results suggest that the transition state is structurally close to the native state and may be described as a distorted native state that is unfolded, or exposed to solvent, in local regions.

Segawa & Sugihara (132) find that the transition state is close to the native state in heat capacity, and the large ΔCp observed in equilibrium unfolding appears in the temperature dependence of the refolding rate constant. On the other hand, there is a substantial (48 kcal/mol) enthalpy difference between the transition state and the native state, which is independent of temperature. It gives rise to a large, constant, temperature dependence of the unfolding rate constant. They conclude from the ΔCp‡ of refolding that the hydrophobic core of hen lysozyme is largely intact and unsolvated in the transition state, and they attribute the ΔH‡ of the unfolding reaction to breaking interactions other than the hydrophobic interaction in going from the native to the transition state. The dependences on GuHCl concentration of the unfolding and refolding rate constants also provide information about the hydrophobic core, if these dependences reflect primarily the change in solvent-accessible surface area of the protein on going to the transition state. Data for α-LA (130) and a mutant of T4 lysozyme (134) suggest that the hydrophobic core is at least partly intact in the transition state.

The mutant of T4 lysozyme studied by Schellman and coworkers shows both cold-induced and heat-induced unfolding reactions, and they have studied the cold-induced unfolding (134, 136). They find that its transition state has properties like those of heat-induced unfolding in other proteins. Additional evidence that the transition state is structurally close to the native state of α-LA is provided by a study of the effect of Ca^{2+} on the unfolding and refolding reactions. Native α-LA has a strong binding site for a single Ca^{2+}, and this site is present, although not as strong, in the transition state (130).

Site-directed mutagenesis can be used to find out if a given interaction that is present in the native state is also present in the transition state, by using mutations to eliminate the interaction. The background of the method has been developed by Matthews and coworkers (137, 138), and a description of the mutational approach to the study of the transition state, as well as results for barnase, has been given recently by Fersht and coworkers (135). The results suggest that the secondary structure of barnase is largely intact in the transition state and that the hydrophobic core is present but weakened.

Proline Isomerization

In 1982, proline isomerization was widely accepted as the plausible explanation for two or more forms of an unfolded protein, but there was no direct

proof. Much of the work aimed at testing the proline hypothesis had been done with RNaseA, the first protein found to have both slow-folding (U_S) and fast-folding (U_F) forms of the unfolded protein (139). Since then, the relation of proline isomerization to folding kinetics has been studied in several proteins, notably hen lysozyme (140), a mouse light-chain immunoglobulin (141), *E. coli* thioredoxin (142, 143), yeast iso-1 (144) and iso-2 (145, 146) cyt c, RNaseT1 (147–149), and porcine RNase (150). Moreover, three new approaches have been introduced with the aim of providing a direct test of the proline hypothesis: isomer-specific proteolysis (151, 152), site-directed mutagenesis (143–147), and enzymatic catalysis of proline isomerization (153). Also, another new approach has been found by accident: NMR detection of proline *cis-trans* isomerization in folded proteins (154–156).

The 1975 proline isomerism model of Brandts and coworkers (157) has two parts combined into a unitary model. The unfolding part explains the existence of both U_F and U_S forms of an unfolded protein, and the refolding part explains why a U_S species is slow folding. The unitary model is: $N \rightleftharpoons U_F \rightleftharpoons U_S$, where the slow $U_F \rightleftharpoons U_S$ reaction takes place only in the unfolded protein and is *cis-trans* isomerization about one or more X-Pro peptide bonds. If more than one proline residue is involved, then at least two U_S species are present. The unfolding part of the model has not been challenged, but the refolding part has been found wanting, because proline isomerization during refolding can also take place in folding intermediates or even after folding is basically complete. The major U_S species of RNaseA refolds to a nativelike intermediate I_N (125, 126, 158) before reversal of isomerization occurs in the final step of folding, $I_N \rightleftharpoons N$. Remarkably, I_N is not only folded but also catalytically active (125). Reversal of isomerization during folding is monitored via unfolding assays (158): I_N unfolds to U_S, whereas N unfolds to U_F. That proline isomerization can occur after folding has been confirmed by NMR studies of SNase (154, 155) and calbindin (156).

Several approaches have been used to test the proline hypothesis. First, the kinetic properties of proline isomerization have been compared with those of the $U_F \rightleftharpoons U_S$ reaction in unfolded RNaseA (159–161). The most distinctive property is acid catalysis of the $U_F \rightleftharpoons U_S$ reaction (159), which takes place only in strong acid ($HClO_4 \geq 3M$), as expected for proline isomerization.

Isomer-specific proteolysis (151, 152) was the second approach used. Proline-specific dipeptidases are known that cleave an X-Pro bond only if the proline residue is *trans* (151), and a similar approach was found possible with trypsin and chymotrypsin (152). When applied to RNaseA, the method gave the surprising result that neither of the two *cis* proline residues (Pro-93, Pro-114) apparently isomerizes sufficiently to *trans* after unfolding to explain the large proportion of U_S^{II}, the major slow-folding species in unfolded RNaseA (162, 163).

A third approach is to compare the refolding kinetics of homologous proteins from different species, or other sources, to identify a slow phase in refolding that depends on a particular proline residue. This approach is a precursor to site-directed mutagenesis, and it has produced some interesting results. Applied to three carp parvalbumins, one with a single proline residue and the other two with no proline residue, the method gave results consistent with the proline hypothesis: the parvalbumin that contains a proline residue shows a slow phase not seen in the two proline-free proteins (164). Ribonucleases from the red deer, the roe deer, the cow (RNaseA), and the sheep show strikingly similar folding kinetics (165), but pig ribonuclease shows a major new slow-folding reaction (150), which is probably correlated with a new proline residue (Pro-115) next to the existing *cis* residue Pro-114. There is a tyrosine fluorescence change in unfolded RNaseA that has the same rate as the $U_F \rightleftharpoons U_S$ reaction (166), and species comparison studies identify the residue responsible as Tyr-92 (126, 167), which is next to the *cis* residue Pro-93. Studies based on this finding (167) provide evidence that Pro-93 is responsible for the major slow-folding species of unfolded RNaseA, in contrast to the conclusion based on isomer-specific proteolysis (163).

The fourth approach is site-directed mutagenesis. It has been applied to the refolding kinetics of *E. coli* thioredoxin (143), yeast iso-1 (144) and iso-2 (145, 146) cyt c, and RNaseT1 (147). The results show both the possibilities and the limitations of the method. In the simplest case, replacement of the proline residue responsible for forming a given U_S species should cause just the disappearance of the corresponding slow phase in refolding. Such an effect is seen when the *cis* residue Pro-55 is replaced in RNaseT1: the amount of U_F species increases, while the amount of the major U_S species is essentially eliminated (147). In other cases, however, replacement of a proline residue, and especially a *cis* proline, may change the refolding rates of all unfolded species. Substituting the single *cis* residue (Pro-76) of thioredoxin by alanine did cause the disappearance of a slow phase in refolding (143), but it also affected the entire refolding kinetics of the protein. Little change in the kinetics of refolding was observed for replacement of the *trans* residue Pro-71 in yeast iso-1 (144) or iso-2 (145) cyt c. On the other hand, replacement of the conserved *trans* residue Pro-76 in iso-2 did cause the disappearance of a slow phase in refolding (146).

The fifth approach is to find out if prolyl *cis-trans* isomerase (PPIase) can catalyze the slow-folding reactions of proteins. PPIase has been purified from pig kidney (168), by using proline-containing peptides as substrates. It has been found to catalyze the slow-folding reactions of some proteins (153), including a mouse immunoglobulin light chain, S-protein (residues 21–124 of RNaseA), and porcine RNase, but it does not catalyze the major slow-folding reaction of RNaseA (153, 168). S-protein contains all four proline residues of

RNaseA, and the entire slow-folding reaction of S-protein is catalyzed by PPIase. The explanation (153) probably is that PPIase cannot act on well-structured folding intermediates, just as it cannot act on folded proteins, and these intermediates are present during the refolding of RNaseA but not S-protein. Thus a positive result with PPIase is meaningful, but a negative result is not. PPIase appears to be the same protein as the widely distributed cyclosporin-binding protein cyclophilin (169, 170), which is thought to function in immune suppression. Cyclosporin (a cyclic peptide with a *cis* Leu-Leu peptide bond) is a potent inhibitor of isomerase activity. The connection, if any, between catalysis of proline isomerization and immune suppression is not yet clear.

The last approach, which involves refolding kinetics less directly, is to detect proline isomerization in a folded protein by NMR. This has been done both for SNase (154, 155) and calbindin (156). The results demonstrate that proline isomerization can occur fairly rapidly in a folded protein, and they also show that two conformations of a protein that differ by a *cis* versus a *trans* proline isomer can be similar enough in stability to coexist in a measurable ratio.

ENERGETICS OF FOLDING

This subject deserves its own review. Many papers have been written and much has been learned in the last few years. We have space only to give a brief summary, with emphasis on developments that concern folding intermediates. A short list of main points is as follows. (*a*) Stripping H_2O from nonpolar side chains to form a hydrophobic core provides the main source of free energy stabilizing a folded protein. This concept, put forth by Kauzmann (171) in 1959, is now accepted by nearly everyone. The alternatives have been scrutinized carefully, as techniques have been developed for measuring individual effects that contribute to protein stability, and the consensus is that the hydrophobic interaction is dominant. (*b*) Isolated secondary structures, in particular α-helices, have been found to be surprisingly stable in some cases. The underlying reasons are only beginning to be understood in energetic terms, but the significance of this development for understanding folding intermediates is unmistakable. (*c*) Nonpolar side chains interact significantly via the hydrophobic effect even when partly or wholly surrounded by H_2O. This subject is not well understood at present, but it is important for understanding the energetics of early folding intermediates. (*d*) A single strong interaction, such as a disulfide bond, can markedly affect the stability of an entire folding intermediate. (*e*) Protein stability may depend strongly on packing side chains in the hydrophobic core in preferred rotamer conformations. (*f*) Measurements can now be made of the strength of individual

interactions, such as salt bridges and charged group-helix dipole interactions, that contribute to protein stability.

The Hydrophobic Interaction

Although the importance of the hydrophobic interaction in protein folding has become better understood, the nature of the hydrophobic interaction has become more complicated. This can be illustrated by theoretical and experimental studies of the pairwise interaction of two nonpolar molecules in H_2O. If the two interacting molecules are small, such as two CH_4 molecules, then the solvent-separated dimer conformation is found to be more stable than the contact conformation (93, 94), and the same result is found theoretically for a CH_4 molecule adjacent to a paraffin wall (95). This finding challenges the model, beloved by biochemists for its simplicity, that the strength of the hydrophobic interaction is proportional to the amount of nonpolar surface buried when the interaction is made. If the simple model held in this case, the contact conformation would be more stable. A further challenge is presented by the finding (96) that even large nonpolar molecules such as cyclohexane interact to form a dimer more weakly in H_2O than in the gas phase. Again, this result contradicts the intuitive picture of the hydrophobic interaction in which H_2O drives two nonpolar molecules together, in order to reduce their contact with the H_2O structure.

The other side of the coin is shown by studies of the hydrophobic collapse of denatured proteins. In particular, studies of C-terminal deletions of SNase and calorimetric investigation of the collapsed form in α-LA (see section on NATURE OF FOLDING INTERMEDIATES) demonstrate that a collapse of the denatured protein occurs at low denaturant concentrations. Mutational substitution of hydrophobic side chains has a major effect on the stability of the collapsed form, and the process may involve hydrophobic interactions between solvated side chains.

Quantitation of the hydrophobic interaction still rests on model compound studies and on the calibration between ΔG° and buried nonpolar surface area (172). Much work has been done both to improve our understanding of the model compound studies per se and to test the validity of small molecule studies as models for the hydrophobic interaction in proteins. Measurements of free energy of transfer on small molecules still provide the basic data for quantitating the bulk hydrophobic interaction in protein folding. As in the pioneering work by Nozaki & Tanford (173), the bulk interaction is measured from the relative solubilities of a model compound in water and in a nonpolar phase, but there is little agreement at present as to what the nonpolar phase should be, in order to represent best the nonpolar interior of a folded protein.

If the model compounds are liquid hydrocarbons, then transfer from H_2O to the neat hydrocarbon liquid can be studied readily and the accurate calorimet-

ric data of Gill and coworkers (174) can be used to analyze the temperature dependence of the hydrophobic interaction (14, 175). A connection has been made between model compound studies and measurements of protein unfolding by analyzing Privalov's "magic temperature" (14, 175) and also by analyzing the relation between ΔCp and buried nonpolar surface area (176) or between ΔCp and ΔS° (177), where ΔS° is the entropy of transfer. Some basic facts have emerged from these studies. The proportionality constant that relates ΔCp to buried nonpolar surface area has been found to be the same in protein unfolding as in model compound studies (176) and the same is true of the constant relating ΔCp to ΔS° (177). On the other hand, the constant relating ΔG° to buried nonpolar surface area (176, 178, 179) is currently in question. Although workers in the protein folding field usually cite model compound data for transfer from H_2O to H_2O-saturated n-octanol (178), Radzicka & Wolfenden (179) have shown that the free energies of transfer to the nonpolar solvent cyclohexane are more than twice as large. Increasing attention is being paid to the suggestion (180) that data for transfer to the gas phase should be put to use, even though the hydrophobic interior of a protein cannot be modeled as a vacuum. Fersht and coworkers (181, 182) have discussed how data for transfer to the gas phase may be used to model the effect on the unfolded protein of mutating one hydrophobic side chain to another.

A new way of thinking about the interaction of H_2O with a protein during the folding process has been put forward by Ooi and coworkers (183, 184). Since H_2O interacts more strongly with polar than with nonpolar groups, they consider the interaction of H_2O with all atoms of the protein during folding. The folding reaction in H_2O is treated as one side of a four-sided cycle; the other sides are folding in vacuo and transfer of the folded and unfolded forms from H_2O to the gas phase. The transfer reactions are modeled from model compound data, using atomic hydrophobicities and H_2O-accessible surface areas, computed from the Lee & Richards algorithm (185). Unknown quantities in the cycle are eliminated by an ingenious use of thermodynamic relations valid at T_m. Two strong points are that they avoid the use of any nonpolar solvent as a reference phase and they take explicit account of polar groups that are buried during folding. Burial of polar atoms during folding is taken into account by Eisenberg & McLachlan (185a) by using atomic hydrophobicities based on the transfer of model compounds between H_2O and H_2O-saturated n-octanol.

Other Interactions

Space permits only a brief mention of current work on other interactions in protein folding. A useful approach to protein stability that considers the

effective concentrations of interacting groups within the protein has been given by Creighton (47).

For a long time, the peptide H-bond was regarded as energetically negligible, because the peptide NH and CO groups can make competing H-bonds to H_2O that are of approximately equal strength. There are, however, some indications that peptide H-bonds are energetically significant and, because there are so many H-bonds in a protein, even a small contribution per H-bond becomes important when multiplied by the total number of H-bonds. In addition, the strength of any intrinsically weak interaction will depend on context: different H-bonds will have different effective concentrations in native proteins (47). Short, 16-residue alanine peptides, solubilized by the insertion of three lysine residues, form stable helices (186). This finding suggests strongly that the polypeptide backbone of the α-helix is an intrinsically stable structure in H_2O. Moreover, the thermal unfolding curves of these helices, like those of other short peptides studied earlier, show that helix formation is enthalpy driven, suggesting that the peptide H-bond contributes a favorable enthalpy to helix formation. A new model system (poly-S-carboxymethyl-cysteine) has been developed for studying the β-structure to flexible chain transition (187). Synthetic templates that can be used to make α-helices or β-sheets have been developed by Kemp and coworkers (188, 189) in order to investigate the energetics of secondary structure formation.

Specific interactions in a folded protein can now be isolated by the combined use of site-directed mutagenesis, measurement of the equilibrium constant for a two-state ($U \rightleftharpoons N$) folding reaction, and structure determination by X-ray or NMR methods. This important advance is neatly illustrated by recent studies of a charged group-helix dipole interaction from Fersht's laboratory (190) and from Matthews's laboratory (191) and by the finding of Dahlquist and coworkers (192) that a single salt bridge contributes 3–5 kcal/mol to the stability of T4 lysozyme. Specific side-chain interactions can also be isolated and studied in single helices as shown by an analysis of charged group-helix dipole interactions in the C-peptide helix (193) and of salt bridge interactions in alanine-based peptide helices (194). Analysis of the residues found at the ends of helices suggest that side-chain H-bonds may be energetically important in capping helices in proteins (195, 196).

It has been known since 1978 (197) that amino acid side chains adopt strongly preferred rotamer conformations in proteins (198). The energetic significance of this finding began to emerge as Ponder & Richards (109) and Lim & Sauer (3) analyzed possible ways of close packing the side chains in the hydrophobic cores of proteins. Recently, experiments aimed at testing the energetic effects of including less preferred rotamer conformations inside the hydrophobic core have been designed (199, 200), and much remains to be done.

CONCLUDING REMARKS

The prospect of describing the mechanism of folding for a few small, single-domain proteins appears good. Several methods have been found for circumventing the high cooperativity of protein folding reactions so that intermediates can be characterized in structural terms. The initial results indicate that structures in folding intermediates are nativelike: if this proves to be general, then understanding structural hierarchies in proteins will be an essential component to solution of the protein folding problem. Folding intermediates can often be placed on a pathway, and it still appears that the folding process is basically sequential. Much remains to be discovered about the nature of adhesion reactions that separate these intermediates, but site-directed mutagenesis and improved structural methodologies seem well suited for this purpose. Another puzzle is the relationship between hydrophobic collapse and the process of protein folding. Studies aimed at understanding the energetics of the hydrophobic effect are likely to be a crucial part of solving this puzzle.

ACKNOWLEDGMENTS

We thank Dr. K. Kuwajima for discussion and many colleagues for sharing results and ideas before publication. Work in our laboratories has been supported by the National Institutes of Health (GM41307 to P.S.K. and GM19988 to R.L.B.).

Literature Cited

1. Kim, P. S., Baldwin, R. L. 1982. *Annu. Rev. Biochem.* 51:459–89
2. Alber, T. 1989. *Annu. Rev. Biochem.* 58:765–98
3. Lim, W. A., Sauer, R. T. 1989. *Nature* 339:31–36
4. Lesk, A. M., Chothia, C. 1980. *J. Mol. Biol.* 136:225–70
5. King, J. 1989. *Chem. Eng. News* 67:32–54
6. Jaenicke, R. 1987. *Prog. Biophys. Mol. Biol.* 49:117–237
7. Montelione, G. T., Scheraga, H. A. 1989. *Acc. Chem. Res.* 22:70–76
8. Ptitsyn, O. B. 1987. *J. Protein Chem.* 6:272–93
9. Lapanje, S. 1978. *Physicochemical Aspects of Protein Denaturation*. New York: Wiley
10. Ghélis, C., Yon, J. 1982. *Protein Folding*. New York: Academic
11. Creighton, T. E. 1983. *Proteins: Structures and Molecular Properties*. New York: Freeman. 515 pp.
12. Privalov, P. L. 1979. *Adv. Protein Chem.* 33:167–241
13. Privalov, P. L. 1982. *Adv. Protein Chem.* 35:1–104
14. Privalov, P. L., Gill, S. J. 1988. *Adv. Protein Chem.* 39:191–234
15. Richardson, J. S., Richardson, D. C. 1989. *Trends Biochem. Sci.* 14:304–9
16. DeGrado, W. F. 1988. *Adv. Protein Chem.* 39:51–124
17. Schellman, J. A. 1987. *Annu. Rev. Biophys. Biophys. Chem.* 16:115–37
18. Goldenberg, D. P. 1988. *Annu. Rev. Biophys. Biophys. Chem.* 17:481–507
19. Chothia, C. 1984. *Annu. Rev. Biochem.* 53:537–72
20. Gō, N. 1983. *Annu. Rev. Biophys. Bioeng.* 12:183–210
21. Kuwajima, K. 1989. *Proteins: Struct. Funct. Genet.* 6:87–103
22. Baldwin, R. L. 1989. *Trends Biochem. Sci.* 14:291–94
23. Wright, P. E., Dyson, H. J., Lerner, R. A. 1988. *Biochemistry* 27:7167–75

24. Creighton, T. E. 1985. *J. Phys. Chem.* 89:2452–59
25. Jaenicke, R., ed. 1980. *Protein Folding.* Amsterdam: Elsevier/North Holland Biomed. Press. 587 pp.
26. Randall, L. L., Hardy, S. J. S., Thom, J. R. 1987. *Annu. Rev. Microbiol.* 41:507–41
27. Rothman, J. E. 1989. *Cell* 59:591–601
28. Freedman, R. B. 1989. *Cell* 57:1069–72
29. Kane, J. F., Hartley, D. L. 1988. *Trends Biotechnol.* 6:95–101
30. Marston, F. A. O. 1986. *Biochem. J.* 240:1–12
31. Taylor, W. R. 1988. *Protein Eng.* 2:77–86
32. Creighton, T. E. 1980. *J. Mol. Biol.* 137:61–80
33. Schmid, F. X., Baldwin, R. L. 1979. *J. Mol. Biol.* 135:199–215
34. Kim, P. S., Baldwin, R. L. 1980. *Biochemistry* 19:6124–29
35. Labhardt, A. M. 1984. *Proc. Natl. Acad. Sci. USA* 81:7674–78
36. Kuwajima, K., Yamaya, H., Miwa, S., Sugai, S., Nagamura, T. 1987. *FEBS Lett.* 221:115–18
37. Ohgushi, M., Wada, A. 1983. *FEBS Lett.* 164:21–24
38. Goto, Y., Fink, A. L. 1989. *Biochemistry* 28:945–52
39. Brems, D. N., Havel, H. A. 1989. *Proteins: Struct. Funct. Genet.* 5:93–95
40. Shoemaker, K. R., Fairman, R., Kim, P. S., York, E. J., Stewart, J. M., Baldwin, R. L. 1987. *Cold Spring Harbor Symp. Quant. Biol.* 52:391–98
41. Dyson, H. J., Rance, M., Houghten, R. A., Lerner, R. A., Wright, P. E. 1988. *J. Mol. Biol.* 201:161–200
42. Oas, T. G., Kim, P. S. 1988. *Nature* 336:42–48
43. Staley, J. P., Kim, P. S. 1990. *Nature.* In press
44. Udgaonkar, J. B., Baldwin, R. L. 1988. *Nature* 335:694–99
45. Roder, H., Elove, G. A., Englander, S. W. 1988. *Nature* 335:700–4
46. Creighton, T. E. 1975. *J. Mol. Biol.* 95:167–99
47. Creighton, T. E. 1983. *Biopolymers* 22:49–58
48. Creighton, T. E., Goldenberg, D. P. 1984. *J. Mol. Biol.* 179:497–526
49. Creighton, T. E. 1978. *Prog. Biophys. Mol. Biol.* 33:231–97
50. Creighton, T. E. 1974. *J. Mol. Biol.* 87:603–24
51. Creighton, T. E. 1977. *J. Mol. Biol.* 113:275–93
52. Marks, C. B., Naderi, H., Kosen, P. A., Kuntz, I. D., Anderson, S. 1987. *Science* 235:1370–73
53. Goldenberg, D. P. 1988. *Biochemistry* 27:2481–89
54. States, D. J., Dobson, C. M., Karplus, M., Creighton, T. E. 1984. *J. Mol. Biol.* 174:411–18
55. Eigenbrot, C., Randal, M., Kossiakoff, A. A. 1990. Submitted
56. Goldenberg, D. P., Creighton, T. E. 1984. *J. Mol. Biol.* 179:527–45
57. States, D. J., Creighton, T. E., Dobson, C. M., Karplus, M. 1987. *J. Mol. Biol.* 195:731–39
58. Deleted in proof
59. Wüthrich, K., Wagner, G. 1978. *Trends Biochem. Sci.* 3:227–30
60. Brown, J. E., Klee, W. A. 1971. *Biochemistry* 10:470–76
61. Bierzynski, A., Kim, P. S., Baldwin, R. L. 1982. *Proc. Natl. Acad. Sci. USA* 79:2470–74
62. Kim, P. S., Baldwin, R. L. 1984. *Nature* 307:329–34
63. Goodman, E. M., Kim, P. S. 1989. *Biochemistry* 28:4343–47
64. Dyson, H. J., Rance, M., Houghten, R. A., Wright, P. E., Lerner, R. A. 1988. *J. Mol. Biol.* 201:201–17
65. Osterhout, J. J. Jr., Baldwin, R. L., York, E. J., Stewart, J. M., Dyson, H. J., Wright, P. E. 1989. *Biochemistry* 28:7059–64
66. Rico, M., Gallego, E., Santoro, J., Bermejo, F. J., Nieto, J. L., Herranz, J. 1984. *Biochem. Biophys. Res. Commun.* 123:757–63
67. Wlodawer, A., Sjölin, L. 1983. *Biochemistry* 22:2720–28
68. Rico, M., Nieto, J. L., Santoro, J., Bermejo, F. J., Herranz, J., Gallego, E. 1983. *FEBS Lett.* 162:314–19
69. Nelson, J. W., Kallenbach, N. R. 1989. *Biochemistry* 28:5256–61
70. Pease, J. H. B., Storrs, R. W., Wemmer, D. E. 1990. Submitted
71. Alber, T., Bell, J. A., Dao-Pin, S., Nicholson, H., Wozniak, J. A., et al. 1988. *Science* 239:631–35
72. Hall, J. G., Frieden, C. 1989. *Proc. Natl. Acad. Sci. USA* 86:3060–64
73. Dalzoppo, D., Vita, C., Fontana, A. 1985. *J. Mol. Biol.* 182:331–40
74. Vita, C., Fontana, A., Jaenicke, R. 1989. *Eur. J. Biochem.* 183:513–18
75. Dyson, H. J., Cross, K. J., Houghten, R. A., Wilson, I. A., Wright, P. E., Lerner, R. A. 1985. *Nature* 318:480–83
76. Wilson, I. A., Skehel, J. J., Wiley, D. C. 1981. *Nature* 289:336–73
77. Udgaonkar, J. B., Baldwin, R. L. 1989. *Nature.* Submitted
78. Roder, H. 1989. *Methods Enzymol.* 176:446–73
79. Kuwajima, K., Nitta, K., Yoneyama,

M., Sugai, S. 1976. *J. Mol. Biol.* 106:359–73
80. Kuwajima, K. 1977. *J. Mol. Biol.* 114:241–58
81. Dolgikh, D. A., Gilmanshin, R. I., Brazhnikov, E. V., Bychkova, V. E., Semisotnov, G. V., et al. 1981. *FEBS Lett.* 136:311–15
82. Gast, K., Zirwer, D., Welfle, H., Bychkova, V. E., Ptitsyn, O. B. 1986. *Int. J. Biol. Macromol.* 8:231–36
83. Pfeil, W. 1981. *Biophys. Chem.* 13:181–86
84. Pfeil, W., Bychkova, V. E., Ptitsyn, O. B. 1986. *FEBS Lett.* 198:287–91
85. Pfeil, W. 1987. *Biochim. Biophys. Acta* 911:114–16
86. Privalov, P. L., Tiktopulo, E. I., Venyaminov, S. Y., Griko, Y. V., Makhatadze, G. I., Khechinashvili, N. N. 1989. *J. Mol. Biol.* 205:737–50
87. Baum, J., Dobson, C. M., Evans, P. A., Hanley, C. 1989. *Biochemistry* 28:7–13
88. Deleted in proof
89. Deleted in proof
90. Dill, K. A. 1985. *Biochemistry* 24: 1501–9
91. Chan, H. S., Dill, K. A. 1989. Submitted
92. Bowie, J. U., Sauer, R. T. 1989. *Proc. Natl. Acad. Sci. USA* 86:2152–56
93. Pratt, L. R., Chandler, D. 1977. *J. Chem. Phys.* 67:3683–704
94. Pangali, C., Rao, M., Berne, B. J. 1979. *J. Chem. Phys.* 71:2975–81
95. Wallqvist, A., Berne, B. J. 1988. *Chem. Phys. Lett.* 145:26–32
96. Wood, R. H., Thompson, P. T. 1990. *Proc. Natl. Acad. Sci. USA.* 87:946–49
97. Amir, D., Haas, E. 1987. *Biochemistry* 26:2162–75
98. Amir, D., Haas, E. 1988. *Biochemistry* 27:8889–93
99. Creighton, T. E. 1975. *J. Mol. Biol.* 96:777–82
100. Kosen, P. A., Creighton, T. E., Blout, E. R. 1981. *Biochemistry* 20:5744–54
101. Shortle, D., Meeker, A. K. 1989. *Biochemistry* 28:936–44
102. Garvey, E. P., Matthews, C. R. 1989. *Proteins: Struct. Funct. Genet.* 6:259–66
103. Wagner, G. 1983. *Q. Rev. Biophys.* 16:1–57
104. Englander, S. W., Kallenbach, N. R. 1983. *Q. Rev. Biophys.* 16:521–655
105. Taniuchi, H., Parr, G. R., Juillerat, M. A. 1986. *Methods Enzymol.* 131:185–217
106. Kuwajima, K., Baldwin, R. L. 1983. *J. Mol. Biol.* 169:299–323
107. Brems, D. N., Baldwin, R. L. 1984. *J. Mol. Biol.* 180:1141–56
108. Kabsch, W., Sander, C. 1984. *Proc. Natl. Acad. Sci. USA* 81:1075–78
109. Ponder, J. W., Richards, F. M. 1987. *J. Mol. Biol.* 193:775–91
110. Goto, Y., Hamaguchi, K. 1981. *J. Mol. Biol.* 146:321–40
111. Chrunyk, B. A., Matthews, C. R. 1990. *Biochemistry.* 29:2149–54
112. Blacklow, S. C., Raines, R. T., Lim, W. A., Zamore, P. D., Knowles, J. R. 1988. *Biochemistry* 27:1158–67
113. Hughson, F. M., Baldwin, R. L. 1989. *Biochemistry* 28:4415–22
114. Reidhaar-Olson, J. F., Sauer, R. T. 1988. *Science* 241:53–57
115. O'Shea, E. K., Rutkowski, R., Stafford, W. F. III, Kim, P. S. 1989. *Science* 245:646–48
116. Anfinsen, C. B., Haber, E., Sela, M., White, F. H. 1961. *Proc. Natl. Acad. Sci. USA* 47:1309–14
117. Anfinsen, C. B. 1973. *Science* 181:223–30
118. Widom, J., Baldwin, R. L. 1983. *Biopolymers* 22:1595–620
119. Ikai, A., Tanford, C. 1973. *J. Mol. Biol.* 73:145–64
120. Ikai, A., Fish, W. F., Tanford, C. 1973. *J. Mol. Biol.* 73:167–84
121. Tanford, C., Aune, K. C., Ikai, A. 1973. *J. Mol. Biol.* 73:185–97
122. Ikemura, H., Inouye, M. 1988. *J. Biol. Chem.* 263:12959–63
123. Zhu, X., Ohta, Y., Jordan, F., Inouye, M. 1989. *Nature* 339:483–84
123a. Silen, J. L., Agard, D. A. 1989. *Nature* 341:462–64
124. Nall, B. T. 1986. *Biochemistry* 25: 2974–78
125. Schmid, F. X., Blaschek, H. 1981. *Eur. J. Biochem.* 114:111–17
126. Schmid, F. X. 1983. *Biochemistry* 22:4690–96
127. Harrison, S. C., Durbin, R. 1985. *Proc. Natl. Acad. Sci. USA* 82:4028–30
128. Kramers, H. A. 1940. *Physica* 7:285–304
129. Goldenberg, D. P., Frieden, R. W., Haack, J. A., Morrison, T. B. 1989. *Nature* 338:127–32
130. Kuwajima, K., Mitani, M., Sugai, S. 1989. *J. Mol. Biol.* 206:547–61
131. Goldenberg, D. P., Creighton, T. E. 1985. *Biopolymers* 24:167–82
132. Segawa, S., Sugihara, M. 1984. *Biopolymers* 23:2473–88
133. Segawa, S., Sugihara, M. 1984. *Biopolymers* 23:2489–98
134. Chen, B., Baase, W. A., Schellman, J. A. 1989. *Biochemistry* 28:691–99
135. Matouschek, A., Kellis, J. T. Jr., Serrano, L., Fersht, A. R. 1989. *Nature* 340:122–26

136. Chen, B., Schellman, J. A. 1989. *Biochemistry* 28:685–91
137. Matthews, C. R. 1987. *Methods Enzymol.* 154:498–511
138. Garvey, E. P., Matthews, C. R. 1989. *Biochemistry* 28:2083–93
139. Garel, J.-R., Baldwin, R. L. 1973. *Proc. Natl. Acad. Sci. USA* 70:3347–51
140. Kato, S., Shimamoto, N., Utiyama, H. 1982. *Biochemistry* 21:38–43
141. Goto, Y., Hamaguchi, K. 1982. *J. Mol. Biol.* 156:891–910
142. Kelley, R. F., Stellwagen, E. 1984. *Biochemistry* 23:5095–102
143. Kelley, R. F., Richards, F. M. 1987. *Biochemistry* 26:6765–74
144. Ramdas, L., Nall, B. T. 1986. *Biochemistry* 25:6959–64
145. White, T. B., Berget, P. B., Nall, B. T. 1987. *Biochemistry* 26:4358–66
146. Wood, L. C., White, T. B., Ramdas, L., Nall, B. T. 1988. *Biochemistry* 27:8562–68
147. Kiefhaber, T., Grunert, H.-P., Hahn, U., Schmid, F. X. 1989. Submitted
148. Kiefhaber, T., Quass, R., Hahn, U., Schmid, F. X. 1989. *Biochemistry*. Submitted
149. Kiefhaber, T., Quass, R., Hahn, U., Schmid, F.X. 1989. *Biochemistry*. Submitted
150. Grafl, R., Lang, K., Wrba, A., Schmid, F. X. 1986. *J. Mol. Biol.* 191:281–93
151. Lin, L.-N., Brandts, J. F. 1979. *Biochemistry* 18:43–47
152. Lin, L.-N., Brandts, J. F. 1985. *Biochemistry* 24:6533–38
153. Lang, K., Schmid, F. X., Fischer, G. 1987. *Nature* 329:268–70
154. Evans, P. A., Dobson, C. M., Kautz, R. A., Hatfull, G., Fox, R. O. 1987. *Nature* 329:266–68
155. Evans, P. A., Kautz, R. A., Fox, R. ●., Dobson, C. M. 1989. *Biochemistry* 28:362–70
156. Chazin, W. J., Kördel, J., Drakenberg, T., Thulin, E., Brodin, P., et al. 1989. *Proc. Natl. Acad. Sci. USA* 86:2195–98
157. Brandts, J. F., Halvorson, H. R., Brennan, M. 1975. *Biochemistry* 14:4953–63
158. Cook, K. H., Schmid, F. X., Baldwin, R. L. 1979. *Proc. Natl. Acad. Sci. USA* 76:6157–61
159. Schmid, F. X., Baldwin, R. L. 1978. *Proc. Natl. Acad. Sci. USA* 75:4764–68
160. Schmid, F. X., Baldwin, R. L. 1979. *J. Mol. Biol.* 133:285–87
161. Schmid, F. X., Buonocore, M. H., Baldwin, R. L. 1984. *Biochemistry* 23: 3389–94
162. Lin, L.-N., Brandts, J. F. 1984. *Biochemistry* 23:5713–23
163. Lin, L.-N., Brandts, J. F. 1987. *Biochemistry* 26:3537–43
164. Lin, L.-N., Brandts, J. F. 1978. *Biochemistry* 17:4102–10
165. Krebs, H., Schmid, F. X., Jaenicke, R. 1985. *Biochemistry* 24:3846–52
166. Rehage, A., Schmid, F. X. 1982. *Biochemistry* 21:1499–505
167. Schmid, F. X., Grafl, R., Wrba, A., Beintema, J. J. 1986. *Proc. Natl. Acad. Sci. USA* 83:872–76
168. Fischer, G., Bang, H. 1985. *Biochim. Biophys. Acta* 828:39–42
169. Fischer, G., Wittmann-Liebold, B., Lang, K., Kiefhaber, T., Schmid, F. X. 1989. *Nature* 337:476–78
170. Takahashi, N., Hayano, T., Suzuki, M. 1989. *Nature* 337:473–75
171. Kauzmann, W. 1959. *Adv. Protein Chem.* 14:1–63
172. Chothia, C. H. 1974. *Nature* 248:338–39
173. Nozaki, Y., Tanford, C. 1971. *J. Biol. Chem.* 246:2211–17
174. Gill, S. J., Nichols, N. F., Wadsö, I. 1976. *J. Chem. Thermodyn.* 8:445–52
175. Baldwin, R. L. 1986. *Proc. Natl. Acad. Sci. USA* 83:8069–72
176. Spolar, R. S., Ha, J.-H., Record, M. T. Jr. 1989. *Proc. Natl. Acad. Sci. USA* 86:8382–85
177. Murphy, K. P., Privalov, P. L., Gill, S. J. 1990. *Science.* 247:559–61
178. Fauchère, J.-L., Pliska, V. 1983. *Eur. J. Med. Chem.-Chim. Therm.* 18:369–75
179. Radzicka, A., Wolfenden, R. 1988. *Biochemistry* 27:1664–70
180. Wolfenden, R., Andersson, L., Cullis, P. M., Southgate, C. C. B. 1981. *Biochemistry* 20:849–55
181. Kellis, J. T. Jr., Nyberg, K., Sali, D., Fersht, A. R. 1988. *Nature* 333:784–86
182. Kellis, J. T. Jr., Nyberg, K., Fersht, A. R. 1989. *Biochemistry* 28:4914–22
183. Ooi, T., Oobatake, M. 1988. *Comments Mol. Cell. Biophys.* 5:233–51
184. Oobatake, M., Ooi, T. 1988. *J. Biochem. Jpn.* 104:433–39
185. Lee, B.-K., Richards, F. M. 1971. *J. Mol. Biol.* 55:379–400
185a. Eisenberg, D., McLachlan, A. D. 1986. *Nature* 319:199–203
186. Marqusee, S., Robbins, V. H., Baldwin, R. L. 1989. *Proc. Natl. Acad. Sci. USA* 86:5286–90
187. Fukada, K., Maeda, H., Ideda, S. 1989. *Macromolecules* 22:640–45
188. Kemp, D. S., Bowen, B. R. 1990. In *The Protein Folding Problem,* ed. J. King, L. Gierasch. Washington, DC: Am. Assoc. Adv. Sci. In press
189. Kemp, D. S., Boyd, J. G., Curran, T.

P., Fotouhi, N. 1990. In *Proc. 11th Am. Peptide Symp.*, ed. J. Rivier, G. Marshall. The Netherlands: ESCOM Sci. Publishers. In press
190. Sali, D., Bycroft, M., Fersht, A. R. 1988. *Nature* 335:740–43
191. Nicholson, H., Becktel, W. J., Matthews, B. W. 1988. *Nature* 336:651–56
192. Anderson, D. E., Becktel, W. J., Dahlquist, F. W. 1989. *Biochemistry*. Submitted
193. Shoemaker, K. R., Kim, P. S., York, E. J., Stewart, J. M., Baldwin, R. L. 1987. *Nature* 326:563–67
194. Marqusee, S., Baldwin, R. L. 1987. *Proc. Natl. Acad. Sci. USA* 84:8898–902
195. Presta, L. G., Rose, G. D. 1988. *Science* 240:1632–41
196. Richardson, J. S., Richardson, D. C. 1988. *Science* 240:1648–52
197. Janin, J., Wodak, S., Levitt, M., Maigret, B. 1978. *J. Mol. Biol.* 125:357–86
198. McGregor, M. J., Islam, S. A., Sternberg, M. J. E. 1987. *J. Mol. Biol.* 198:295–310
199. Sandberg, W. S., Terwilliger, T. C. 1989. *Science* 245:54–57
200. Karpusas, M., Baase, W. A., Matsumura, M., Matthews, B. W. 1989. *Proc. Natl. Acad. Sci. USA* 86:8237–41