Ann. Rev. Biochem. 1982. 51:459–89

SPECIFIC INTERMEDIATES IN THE FOLDING REACTIONS OF SMALL PROTEINS AND THE MECHANISM OF PROTEIN FOLDING[1]

Peter S. Kim and Robert L. Baldwin

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

CONTENTS

[1]Abbreviations used for proteins are: BPTI, bovine pancreatic trypsin inhibitor; cyt *c*, horse heart cytochrome *c*; RNase A, bovine pancreatic ribonuclease A; RNase S, a derivative of RNase A cleaved at the peptide bond between residues 20 and 21; S peptide, residues 1–20 of RNase S; S protein, residues 21–124 of RNase S. Other abbreviations are: N, native; U, unfolded; I, intermediate; I_N, quasinative intermediate; U_F and U_S, fast- and slow-folding species, respectively; *k*, rate constant; GuHCl, guanidinium chloride; t_m, temperature at the midpoint of an unfolding transition; ΔC_P, change in heat capacity at constant pressure; θ, mean residue ellipticity. Lysozyme refers to hen egg white lysozyme and myoglobin to sperm whale myoglobin.

0066-4154/82/0701-0459$02.00

PERSPECTIVES AND SUMMARY

The stage is set for determining the pathway of folding of representative small proteins by characterizing the structures of well-populated folding intermediates. Structural intermediates accumulate in kinetic experiments under conditions (especially low temperatures) where the intermediates are stable relative to the unfolded form. The major problem appears to be in tracing the folding pathway for a single unfolded form, because the *cis-trans* isomerization of proline residues about X-Pro peptide bonds often gives multiple unfolded forms, with different rates of refolding. The role of proline isomerization in refolding is beginning to be understood. Covalent intermediates have been trapped and characterized in the refolding process, which accompanies reoxidation of the disulfide bonds for two small proteins. Equilibrium intermediates have been found and characterized for some unusual small proteins, and it appears that the unfolding reactions induced by certain salts or methanol yield equilibrium intermediates even for proteins that normally show highly cooperative folding.

The pathway of folding should reveal the mechanism of folding and help in determining the code by which the amino acid sequence of a protein specifies its tertiary structure. This will aid in engineering changes in protein structure using recombinant DNA techniques. Knowledge of the pathway can show at which stage an amino acid substitution causes a change in folding.

Earlier work searching for equilibrium intermediates with small, single-domain proteins such as RNase A and lysozyme gave chiefly negative results and led to the use of the two-state approximation ($N \rightleftharpoons U$, N = native, U = unfolded), in which folding intermediates are neglected. Populated intermediates were thought to be ruled out by the success of the two-state approximation. However, such highly cooperative folding is found inside the folding transition zone where intermediates are only marginally stable. Folding can be measured kinetically in conditions where intermediates are more stable.

The evidence is now essentially complete that multiple forms of an unfolded protein are produced by the *cis-trans* isomerization of proline residues about peptide bonds after unfolding. Slow-folding (U_S) and fast-folding (U_F) forms of an unfolded protein can be recognized in refolding experiments by the fact that native protein is formed in separate slow ($U_S \rightarrow N$) and fast ($U_F \rightarrow N$) refolding reactions. They also can be recognized in unfolding experiments by refolding assays made during the fast ($N \rightarrow U_F$) and slow ($U_F \rightleftharpoons U_S$) phases of unfolding. The kinetic properties of the $U_F \rightleftharpoons U_S$ reaction, measured during unfolding, match those of proline isomerization in such specific aspects as catalysis by strong acid and cleavage of X–Pro bonds by an enzyme specific for the *trans* proline isomer. Folding in strongly native conditions occurs before proline isomerization and can go almost to completion. An enzymatically active intermediate that still contains a wrong proline isomer is found when RNase A folds at 0°–10°C. Partial folding increases the rate of proline isomerization, possibly because strain is produced in folding intermediates. The unexpectedly low proportion of U_S species in several unfolded proteins suggests that not all proline residues produce slow-folding species.

Information about the folding pathway is preliminary in all cases but, surprisingly, S–S bond formation in BPTI proceeds via obligatory two-disulfide intermediates each having a nonnative S–S bond. The major folding reaction occurs in a single S–S rearrangement. Both kinetic and equilibrium results for folding with S–S bonds intact are consistent with a framework model in which the H-bonded secondary structure is formed early in folding. The tertiary structures of α-lactalbumin, penicillinase, and carbonic anhydrase are disrupted before their secondary structures unfold, in denaturant-induced unfolding. Kinetically, secondary structure is formed early in the folding of RNase A and RNase S, as judged by stopped-flow CD studies and protection of NH protons against exchange with solvent. The kinetic mechanism of folding appears to be sequential folding with defined intermediates. Folding probably proceeds along a pathway determined by the most stable intermediates. Independently folding domains have been demonstrated via fragment isolation for the small proteins ovomucoid and elastase, the α subunit of tryptophan synthase, and for

larger proteins. For multidomain proteins, the pathway of folding involves independent folding of individual domains followed by domain interaction. Mutants blocked kinetically in the folding and assembly of a trimeric protein, the tail spike protein of phage P22, have been demonstrated.

INTRODUCTION

Statement of the Problem

Small, monomeric proteins fold to thermodynamically stable structures, as judged by reversibility of folding. Nevertheless, folding is very rapid in most cases (seconds or less). The number of possible conformations is astronomical. If the pathway is under thermodynamic control, how does the protein find the most stable structure so quickly? If the pathway is under kinetic control, how are incorrectly folded structures avoided?

The amino acid sequence codes for the folding of a protein, but amino acid substitutions (changes in the code) are allowed at almost all residue positions without drastic changes in the folding pattern. X-ray structures have been determined for globins that are related only distantly through evolution and have only a few amino acids in common; yet the "globin fold" is strikingly similar in each case. The qualitative features of folding appear to be the same in horse cyt *c* and yeast cytochrome *c*, even though they differ by 46% in amino acid sequence. The code for folding is not a simple code like the mRNA triplet code.

The complexity of the code probably reflects the complexity of the folding process. Determination of the pathway of folding may be the chief means of solving the code, because the specific interactions can be measured at different stages in folding. This can be illustrated by a current model for folding, according to which α helices and β sheets form at their correct locations in the otherwise unfolded chain. Amino acid substitutions may be allowed because the choice between helix, sheet, or no folding is averaged over several residues, so that one substitution need not tip the balance, or because only a few specific interactions determine the locations of the α helices and β sheets. At the second stage in folding, the model predicts that α helices and β sheets interact via special pairing sites,[2] which are coded by only a few residues, thus allowing substitutions at other residue positions. A decade ago it was difficult to understand how an amino acid substitution could be tolerated in the interior of a protein because the side chains are closely packed and the protein structure was thought to be rigid. Today it is known, especially from NMR studies of tyrosine and phenylalanine ring flips, that the interior of a protein is flexible.

[2]Alternatively, the pairing between α helices and/or β strands may be determined by their relative positions in the chain (1a).

The practical problems in determining the folding pathway are severe. The methods that give structural information about protein folding (e.g. X-ray, NMR, and CD) are intrinsically slow, so that equilibrium intermediates are needed. Moreover, the intermediates must be well-populated, since these methods require reasonably pure materials. The pathway of folding should first be determined for the simplest case, that of small, "single-domain" proteins like BPTI, RNase A, myoglobin, or staph nuclease. But folding of these small proteins is highly cooperative in most cases, and equilibrium intermediates are not populated. On the other hand, although kinetic intermediates may be well populated, steps in folding are often fast (1–100 msec). It is necessary to find methods of trapping intermediates in a stable form, so that they can be studied at leisure, of slowing down folding, or of adapting spectroscopic methods so that they will give structural information rapidly. Some solutions to these problems have been found and are discussed here.

Other Reviews

An excellent summary of work on protein folding, as of September 1979, is contained in a set of symposium papers collected by Jaenicke (1) into a book, *Protein Folding.* It includes a separate review of recent experimental work. A new monograph on protein folding is being prepared by Ghélis & Yon (2). The study of folding intermediates was reviewed in 1978 by Creighton (3) and in 1975 by Baldwin (4). Wetlaufer (5) has just reviewed the folding of protein fragments. Structural studies of folding, and the possible relationships between the final structure and the pathway of folding have been reviewed recently by Richardson (6), Ptitsyn & Finkelstein (7), Thomas & Shechter (8), and Rossmann & Argos (9). Thermodynamic data on the energetics of folding, obtained by calorimetric studies of folding transitions, have been reviewed by Privalov in 1979 (10). Nemethy & Scheraga consider both theoretical and experimental aspects in a 1977 review (11), which stresses the possible stereochemical determinants of folding. The use of packing principles and surface areas in analyzing the folding process was reviewed in 1977 by Richards (12). Ikegami (13) has just reviewed work on interpreting folding transitions by a cluster model. Earlier general reviews of folding have been given by Anfinsen & Scheraga in 1975 (14) and by Wetlaufer & Ristow in 1973 (15). The nature of protein folding transitions and of the unfolded state was discussed by Tanford (16, 17) in a pair of classic reviews.

MODELS FOR THE FOLDING PATHWAY

These models are of two kinds: kinetic and structural. Since structural data on folding intermediates are only now starting to be available, both classes

of models still consist of guesses about the folding process. Present structural models are based chiefly on reflection about the X-ray structures of native proteins. The aim of a structural model is to give the actual structures of intermediates as well as their order on the pathway, without being too specific about the factors that control the rate of folding. The aim of a kinetic model is to indicate the dominant intermediates and give the factors that control the rate of folding without being too specific about the structures of the intermediates.

Kinetic Models

1. BIASED RANDOM SEARCH The possibility that proteins might fold by a purely random search of all possible conformations was considered by Levinthal (18) and then dismissed, because the time required for folding would be impossibly long [10^{50} years for a chain of 100 residues (19)]. However, it is possible that a biased random search could occur in a reasonable time. By means of a computer-simulated folding for a lattice model, it is possible to show that the number of possible chain conformations is drastically reduced if only self-avoiding (sterically possible) conformations are allowed (M. Levitt, personal communication, 1978). Levitt also found that a significant fraction of the self-avoiding conformations are fairly compact, and he suggested that rapid folding might begin whenever the unfolded chain assumes a backbone conformation sufficiently like that of the native protein, because the major free energy barrier to folding (the necessary reduction in entropy of the polypeptide chain) has been overcome [compare (20)]. This result, taken together with the possibility of a rapid, nonspecific collapse when refolding is initiated, [compare (21)] suggests that a biased random search could play an important role in early stages of folding. However, it may not be easy to test the prediction that a nonspecific collapse precedes specific folding, because specific structures can be formed very rapidly: α helix formation in model systems occurs in 10^{-5}–10^{-7} sec (22–24).

2. NUCLEATION-GROWTH The term nucleation has been used with two quite different meanings to describe models for protein folding. In both cases the nucleus is the structure formed at the beginning of folding, which guides subsequent steps. With the first meaning of nucleation, folding is sequential, and early folding intermediates may be populated: the nucleated molecule may be directly observable in kinetic folding experiments. With the second meaning, the folding reaction proceeds rapidly as soon as a nucleus is provided (as in crystallization of a supercooled liquid after seeding with a crystal, or as in α helix formation) and the nucleated molecule is not observable as a populated species either because folding occurs rap-

idly after nucleation or because the nucleus is unstable by itself and breaks down if it is not stabilized by further folding. The second meaning of nucleation is the classical one in chemistry, and the first meaning is a special usage that has developed gradually in protein folding work. The term was used originally in its correct (second) sense, but when it became apparent that the "nucleated" molecules might nevertheless be observable in kinetic folding experiments, the term was still retained by several workers. We suggest that the term nucleation now be dropped in protein folding studies unless it is used with its classical meaning, and that the other type of folding be referred to as *sequential folding* in which the first structure formed is the *kernel* (see Model 4).

In the nucleation-growth model, folding cannot start until an initial reaction occurs (nucleation), and subsequent folding takes place rapidly compared to the observed folding reaction. Folding intermediates are not populated, because folding is too fast once it starts, and the rate of folding is determined by the nucleation reaction. Therefore, demonstration of a populated intermediate in folding rules out the nucleation-growth model. Populated kinetic intermediates have been demonstrated in the folding of several proteins (see section on kinetic intermediates), so it appears that protein folding is not a nucleation-limited reaction.

3. DIFFUSION-COLLISION-ADHESION In this *microdomain coalescence* model, short segments of the unfolded chain fold independently into microdomains. These are unstable, but they diffuse, collide, coalesce, and become stable (19, 25). Karplus & Weaver (19, 26) calculated the folding rate for a diffusion-collision model in which the diffusional collision is rate limiting, and computer simulations for the folding of apoMb have been made on the assumption that diffusion is rate limiting (27). However, adhesion or coalescence of the two microdomains may be the rate-limiting step, in this model. The two steps can be written as:

$$\underline{\text{A}\quad\text{B}} \quad \underset{k_{21}}{\overset{k_{12}}{\rightleftharpoons}} \quad \text{A·B} \quad \underset{k_{32}}{\overset{k_{23}}{\rightleftharpoons}} \quad \text{C}$$

where A__B are the two microdomains, linked by the polypeptide chain, which diffuse together to form the encounter complex A·B, and C is the product formed by adhesion. There are two limiting cases: (*a*) If k_{21}, the rate of dissociation of A·B, is large compared to k_{23}, the rate of adhesion, then $v = (k_{12}/k_{21})k_{23}$, where v is the overall rate of forming C from A__B. In this case, it is the equilibrium ratio of A·B to A__B and not the diffusion-controlled rate of forming A·B that enters into the overall rate expression. This situation is found commonly for reactions in solution (28) because k_{21} is always large (of the order of 10^{10} s^{-1}). (*b*) In special cases, such as proton transfer reactions, k_{23} can be very large (10^{12} s^{-1} for proton

transfer) so that $k_{23} >> k_{21}$ and $v = k_{12}$. This is the situation envisaged in the original diffusion-collision model.

The diffusion-collision model (19, 26) has been tested (29, 30) by asking if the folding rate of RNase A depends on solvent viscosity in the direct ($U_F \rightarrow N$) folding reaction. Model compound studies have shown that diffusion of one segment of a chain molecule relative to another is dependent on solvent viscosity (31). The overall folding rate of RNase A was found to be independent of solvent viscosity (29) when either glycerol or sucrose was added, which demonstrates that diffusion is not rate limiting. Recently, a fast reaction (msec) of RNase A has been found (32) that is strongly affected by solvent additives that change the viscosity. Since it can be measured at temperatures far below the transition zone for unfolding, its relation to the folding process is not yet clear.

4. SEQUENTIAL FOLDING Folding occurs in a unique and definite sequence of steps, analogous to a metabolic pathway. Intermediates may be populated in suitable conditions of folding. To demonstrate sequential folding, it is necessary to show that there are specific, well-populated intermediates. This test has been satisfied for the folding at low temperatures (0°–10°C) of the major slow-folding species of RNase A (33–38) and also of RNase S (30, 39, 40). Proline isomerization is one step in these folding reactions, but folding can proceed to an enzymatically active, native-like form of RNase A before proline isomerization occurs (33, 36–38). Therefore, the folding process is probably similar (i.e. sequential) for both the fast and slow folding species.

Structural Models

The increasing number of X-ray crystal structures has stimulated the proposal of many structural models for protein folding. The possible relationship between the folding pathway and the final structure of a protein has been the subject of several recent reviews (6–9, 12, 25, 41–46) and we do not review these models here.

Some models for the folding of an entire class of proteins postulate that folding begins by forming a "primitive" H-bonded structure that breaks down to generate the observed structure. The primitive postulated by Ptitsyn & Finkelstein (7) for all β proteins is a long two-stranded antiparallel β structure with a central hairpin loop. The Greek key pattern of connections between β strands (47) then results from breaking this hairpin helix into shorter segments by opening unpaired loops. The particular "swirl" of the Greek key (only one is found, and not its isomer) results from the right-handed twist of the β sheet. An α-helical folding primitive has been postulated by Lim (45).

WORKING MODELS In comparing experimental results with models, experimentalists have two choices: either to take an existing model and to test it against their results, or else to extract a working model from the experimental data. Since present structural information is "low resolution," such working models are necessarily low resolution. Two generalized working models are being tested currently. The first is the *framework model* in which the H-bonded secondary structure is formed early in folding. The second is *modular assembly,* in which essentially complete folding of any part of a protein occurs at one time, although different parts of the protein fold at different times (folding by parts). Note that the folding process may combine features of both models: formation of H-bonded secondary structure may precede tertiary interactions, as in the framework model, while separate subdomains (each capable of forming its own secondary structure) may fold at different times, as in modular assembly.

PROLINE ISOMERIZATION AND SLOW-FOLDING SPECIES

Formation of Slow-Folding Species

Proline isomerization as a slow step in protein folding was suggested by Brandts and co-workers (48) as a possible explanation for the two unfolded forms of RNase A (49). The folding of RNase A shows biphasic kinetics: a fast phase (50 msec at 25°C) precedes a major (80%) slow phase (20 sec at 25°C); both the fast- and the slow-folding reactions produce native enzyme (49). In unfolding experiments, the fast-folding species is formed rapidly, and at least two slow-folding species (U_S^{II}, U_S^{I}) are formed slowly (38, 48, 50–52). A quantitative study of the unfolding and refolding kinetics of RNase A (53) has demonstrated that the 3-species mechanism

$$N \longleftrightarrows U_F \longleftrightarrows U_S$$

explains the unfolding and refolding kinetics in the folding transition zone, where N is only marginally stable, and folding intermediates are not well populated. The fast-folding species (U_F) of RNase A is not a partly folded or nucleated molecule, since its concentration (20% of the unfolded molecules) is not affected by high temperature or strong denaturants such as 6 M GuHCl or 8.5 M urea (49, 54). Recently, the same tests used for RNase A have demonstrated the existence of both fast- and slow-folding molecules in hen lysozyme (55, 56) and in horse cyt *c* (57). Urea-gradient electrophoresis experiments at 2°C show that unfolded chymotrypsinogen and α-chymotrypsin also contain both slow- and fast-folding molecules

(58). Unfolding occurs in two stages, consistent with a $N \rightarrow U_F \rightleftharpoons U_S$ unfolding mechanism for a BPTI derivative (59), yeast isocyt *c* (60), and pepsinogen (61).

The first good evidence that the $U_F \rightleftharpoons U_S$ reaction of RNase A is proline isomerization was based on a comparison of the kinetics of the $U_F \rightleftharpoons U_S$ reaction in the unfolded protein (51) with the *cis* $\rightleftharpoons$ *trans* isomerization of prolyl residues in model compounds (48, 50, 62–65), and it made use of the conclusion that both U_F and U_S are completely unfolded (49, 54). The most striking characteristics, which are common to both reactions, are (51): (*a*) a high activation enthalpy ($\sim$ 20 kcal/mol); (*b*) catalysis by strong acids; and (*c*) kinetics that are independent of GuHCl concentration, which confirm that the interconversion of U_S and U_F does not involve residual structure (223). RNase A has three nitratable tyrosine groups including Tyr 115, which follows Pro 114; the kinetics of the $U_F \rightleftharpoons U_S$ reaction for nitrotyrosyl RNase A have been correlated with the pK changes during unfolding (66), and it has been suggested that NO_2-Tyr-115 provides an optical probe monitoring isomerization of Pro 114.

In recent work (L. -N. Lin and J. F. Brandts, personal communication, 1981), the appearance of a specific wrong proline isomer during unfolding has been correlated directly with the formation of a slow-folding species. Enzymatic cleavage specific for the *trans* X-Pro bond has been used to break the peptide bond between Tyr 92 and Pro 93 during the unfolding of RNase A, and the results can be correlated with the formation of U_S^{II}. Pro 93 is *cis* in native Rnase A and *trans* in U_S^{II} (the major unfolded species).

Proline Isomerization During Folding

Although the $U_F \rightleftharpoons U_S$ reaction of RNase A is almost certainly proline isomerization, the refolding of U_S^{II} has kinetic properties very different from proline isomerization. The activation enthalpy for the $U_S \rightarrow N$ reaction is small ($<$ 5 kcal/mol at pH 6 and 20-40°C) as compared with 20 kcal/mol for proline isomerization (50). Furthermore, the rate of the $U_S \rightarrow N$ reaction is strongly dependent on the GuHCl concentration, unlike proline isomerization (50). Thus, proline isomerization is not the initial and rate-limiting step in the folding of the major U_S species of RNase A, in contrast to the original proposal (48).

At low temperatures (0°–10°C), a native-like intermediate (I_N) is formed in the folding of U_S^{II}. I_N has nearly the same tyrosine absorbance and enzymatic activity as the native protein (33, 38), but it differs from N in having a wrong proline isomer. I_N unfolds to give U_S, whereas N unfolds to give U_F (33). Proline isomerization appears to be the final or nearly final

step in folding ($I_N \rightarrow N$). Proline isomerization can be 20 to 40 times faster in I_N than in the unfolded protein (33), perhaps because the non-native proline residues are under strain in I_N. The rate of proline isomerization in a cyclic pentapeptide is six times faster than in the corresponding blocked linear peptide, probably because of strain (67). In the case of U^{II}_S, an incorrect proline isomer does not block folding in moderate or strongly native folding conditions, but rather slows down the folding process: the probable explanation is that folding intermediates are less stable with a non-native proline isomer.

The kinetics of folding for three different carp parvalbumins provide further evidence for the role of proline isomerization in protein folding (68, 69). They have similar amino acid sequences and spectroscopic properties, but one of the parvalbumins contains a proline residue and the other two do not (69). All three proteins show complex folding kinetics; however, the parvalbumin that contains a proline residue shows an additional slower phase not seen in the other two proline-free proteins, and this phase has some kinetic properties like those of proline isomerization (69). However, a similar comparison of cytochrome *c* molecules from two different species has not given comparable results (70). The probable existence of nonessential proline residues complicates this kind of comparison. The role of proline isomerization has been studied in the refolding kinetics of a specific fragment of procollagen (71).

Nonessential Proline Residues

Not all prolines in a protein may affect the kinetics of folding: there may be "essential" and "nonessential" proline residues (51). The X-ray crystal structure of RNase S suggests that Pro 114 may be accommodated in either the *cis* or *trans* configuration [H. W. Wyckoff, quoted in (51)]. Levitt (72), used conformational energy calculations to study the effect of non-native proline isomers in BPTI, which has four *trans* proline residues. Proline residues can be classified into three groups, based on the energy difference between the native protein and the minimum energy structure with a wrong proline isomer (72). He suggested that these types of proline residues should produce different types of folding reactions (59, 72). Type I (small energy difference) should not affect the rate of folding, type II (intermediate energy difference) should slow down but not block folding, and type III (large energy difference) should block folding in the manner originally suggested by Brandts and co-workers (48).

So far, only type II prolines have been characterized (59). The "type" of folding reaction will depend on the folding conditions. In the folding of RNase A at 25°C, the activation enthalpy changes from 3 kcal/mol (type

II folding) in 0.1 M GuHCl to 18 kcal/mol (type III folding) in 2 M GuHCl (37, 50).

Wüthrich and co-workers have studied proline-containing linear oligopeptides and shown that the *cis/trans* ratio and the isomerization rate of the X-Pro bond depend on the charge and nature of the amino acid preceding the proline residue (62, 63, 67, 73, 74).

Nonproline peptide bond isomerization may be an important factor in the folding of some proteins. For example, the crystal structure of carboxypeptidase A shows three *cis* peptide bonds that are not N-terminal to prolyl residues (75).

KINETIC INTERMEDIATES

Multiple Unfolded Forms

The existence of multiple unfolded forms of a protein, arising from proline isomerization, is discussed above. It presents a serious problem in working out the kinetic pathway of folding, since the pathway must be studied separately for each unfolded form. In the case of RNase A, it has been possible to study the major slow-folding species U_S^{II} [(80% of the total slow folding species (38)]. In the case of hen lysozyme, it is possible to study the direct folding reaction ($U_F \rightarrow N$) of the species with correct essential proline isomers (55, 56).

The standard test for the presence of a kinetic intermediate is the existence of two kinetic phases: whenever more than one phase is observed, at least three species must be involved. However, the three species could be U_F, U_S, and N, and the two phases could be $U_F \rightarrow N$ and $U_S \rightarrow N$, so that a structural intermediate need not be present.

Tests for Structural Intermediates

The standard test for a structural intermediate is the *kinetic ratio test,* whose application to the problem of protein folding has been discussed (39). If the folding reaction shows different kinetics when measured by two different probes, then a structural intermediate must be present, provided that all unfolded species appear alike when measured by each probe. If two (or more) well-resolved kinetic phases are found, and the two probes change differently in different phases, then there is at least one structural intermediate that can be studied readily. Three probes that are particularly informative are: (*a*) stopped-flow CD (76, 77); (*b*) enzyme activity, measured by combination with specific ligands (38, 39, 56, 78) or measured directly (38, 49); (*c*) protection of NH protons against exchange with solvent (34, 35). The test of *specific combination* between fragments has been applied to the folding of RNase S (39). The principle is that, if combination between S

peptide (residues 1–20) and S protein (residues 21–124) occurs early in the folding of S protein, then a structural intermediate must be present to provide the combining site.

RNase A and RNase S

The present minimal mechanism for the folding of the U_S^{II} species of RNase A at 0°–10°C is:

$$U_S \rightleftharpoons I_1 \rightleftharpoons I_N \rightleftharpoons N$$

but it is probable that additional intermediates are populated between I_1 and I_N. I_1 has been observed by protection of NH protons against exchange with solvent (34, 35). The protected protons are trapped early and remain trapped throughout folding. The average degree of protection in I_1 is at least 100 (pH 7.5, 10°C) (35) for the 50 most protected NH protons of native RNase A. In native proteins, some NH protons are protected by as much as 10^8 (79, 80). I_N is highly folded as measured by tyrosine absorbance or binding of the specific inhibitor 2'-CMP (33, 38), and I_N is enzymatically active (38). Nevertheless, I_N still contains a wrong proline isomer (33) and the $I_N \rightleftharpoons$ reaction can also be followed by a fluorescence change (36, 37).

The unfolding pathway of RNase A contains an additional intermediate I_U (53):

$$N \rightleftharpoons I_U \rightleftharpoons U_F \rightleftharpoons U_S$$

The refolding kinetics of I_U have been measured, using a sequential stopped-flow apparatus (52), and it is known that proline isomerization does not occur freely in I_U (81). Thus far I_U has been studied only in unfolding conditions, and it is not known whether I_U is also populated in refolding experiments.

The refolding kinetics of RNase S are more complex (39, 40, 82) than those of RNase A, but they contain additional information about the role of the S-peptide moiety (residues 1–20) in folding. Recent stopped-flow CD measurements on the folding of RNase S (U_S) show that sequential steps in folding can be resolved: β-sheet formation precedes the S-peptide α-helix formation (A. M. Labhardt, personal communication, 1981). Enzymatic activity, as measured by the ability to bind 2'-CMP, is regained together with the α-helix formation. It is not yet known when proline isomerization occurs.

Lysozyme

The refolding kinetics of hen egg white lysozyme (83) have recently been reinvestigated by Utiyama and co-workers (55, 56). Their results demon-

strate that unfolded lysozyme also exists in a mixture of fast and slow folding species (U_F:U_S ratio of 90:10) [cf Hagerman (84)]. The refolding kinetics outside the transition zone are biphasic, and both phases produce native enzyme (56). The unfolding kinetics are monophasic, and the formation of U_F precedes U_S (56). Evidence for an early intermediate in folding is based on an absorbance change that occurs in the dead time of the stopped-flow instrument (20 msec), after correction for solvent effects (55). The activation enthalpy for the folding of U_S is only 11 kcal/mol (vs 20 kcal/mol for proline isomerization), which suggests that some folding occurs before proline isomerization in U_S (56).

Cytochrome c

The refolding kinetics of horse Fe(III) cyt *c* (85) have been reinvestigated recently (57). As in RNase A and lysozyme, unfolded cyt *c* consists of an equilibrium mixture of U_F and U_S (in a 78:22 ratio), which gives rise to biphasic refolding kinetics, with native enzyme formed in both phases (57). An earlier suggestion that the fast refolding reaction is the formation of an abortive intermediate (85) has been ruled out. Sequential unfolding/refolding ("double jump") experiments demonstrate that the unfolding of cyt *c* produces U_F, which then isomerizes to U_S (57). Similar results have been obtained with yeast iso-2 cyt *c* (60). An intermediate has been identified in the $U_F \rightarrow N$ reaction of cyt *c;* the Soret absorbance change precedes the recovery of the native 695 nm band spectrum. In the $U_S \rightarrow N$ reaction, an ascorbate-reducible intermediate is formed before native enzyme is produced (57). Several kinetic intermediates have been found in unfolding by monitoring heme absorbance (86).

Carbonic Anhydrase

This protein (mol wt 29,000) is about twice as large as the other proteins discussed above. The folding of bovine carbonic anhydrase is much slower than the folding of RNase A, lysozyme, or cyt *c*. Both refolding and unfolding kinetics show multiple phases (78, 87–89), but their relationship to possible multiple forms of the unfolded protein has not yet been investigated (carbonic anhydrase has 20 prolines). Carbonic anhydrase contains Zn^{2+} and the presence or absence of Zn^{2+} strongly affects the refolding kinetics (90, 91). There is evidence for early formation of the H-bonded framework: changes in θ_{222} (secondary structure) precede changes in θ_{270} (tertiary structure) (88). A spin-label study of carbonic anhydrase shows that an intermediate is formed within 0.1 sec after the start of refolding (89). A late intermediate can bind a specific inhibitor, but does not have enzymatic activity (78). Interestingly, the fastest observed phase in the refolding of carbonic anhydrase has a rate that increases with increasing GuHCl concentration, which suggests that there is an early abortive intermediate

in a rapid preequilibrium with the unfolded state (88). However, carbonic anhydrase is known to precipitate easily during attempts at renaturation (92–94), and it is possible that this is aggregation dependent.

Other Proteins

Creighton (58) has recently introduced urea gradient electrophoresis as a method of studying folding intermediates. A linear gradient of urea perpendicular to the direction of migration is used, and the migration pattern is observed as a function of time; the patterns obtained with native and unfolded proteins are compared (58). The temperature of electrophoresis is low (2°C), to decrease the rates of folding and of proline isomerization. Slow-folding (U_S) forms of the unfolded protein have been demonstrated with RNase A, chymotrypsinogen, and α-chymotrypsin (58). Slow-folding species have not been detected in several other small proteins, including lysozyme and cyt *c*. However, this is not surprising, since reactions with half times less than 8 min are too fast to measure by this method, and the fraction of U_S molecules is small in these proteins (55–57). Several examples of compact kinetic intermediates in folding have been demonstrated with this method (58).

Kinetic and urea gradient electrophoresis experiments also demonstrate the existence of at least two unfolded forms in the α subunit of tryptophan synthase, in addition to two rapidly formed intermediates (95, 96). Moreover, binding of a substrate analogue during refolding displays biphasic kinetics; the rates for the two phases are identical with those observed for folding in the absence of the analogue (96). The α subunit can be cleaved proteolytically into two fragments, each of which can fold by itself, but neither one alone can bind the substrate analogue (97). The relationship between these fragments and the two kinetic intermediates of the intact α subunit is not yet known.

The unfolding of apomyoglobin has been studied by stopped-flow CD and by fluorescence (77) as has the folding of the β chain of hemoglobin (98). Both studies suggest the existence of specific interactions apart from helix formation. The helices of apomyoglobin break down more rapidly in unfolding, as judged by CD, than interactions detected by a fluorescence probe (77). In refolding, the β chain interacts rapidly and specifically with the heme, followed by slower helix formation (98).

EQUILIBRIUM INTERMEDIATES

Tests for Intermediates

The equilibrium unfolding transitions of most small, globular proteins are highly cooperative, and the two-state approximation ($N \rightleftharpoons U$) is usually a good working model for equilibrium studies. However, in the past few years, several examples of proteins with populated equilibrium intermedi-

ates have been reported. There are two tests for an equilibrium intermediate based on the use of probes (16) (Figure 1): (*a*) a biphasic transition as measured by a single probe and (*b*) noncoincident transitions as measured by different probes. Either one of these observations is sufficient evidence for an equilibrium intermediate, and they are not mutually exclusive. There is also a calorimetric test for intermediates (99, 100): $\Delta H_{vH} < \Delta H_{cal}$, where ΔH_{cal} is the calorimetrically determined enthalpy of unfolding and ΔH_{vH} is the apparent enthalpy computed from the temperature dependence of the $N \rightleftharpoons U$ equilibrium constant, by the van't Hoff relation. Equilibrium measurements cannot demonstrate that an intermediate is actually on the pathway of folding: it may be an abortive intermediate or an alternative native form. Also, aggregation of an unfolded protein is known to occur inside the unfolding transition zone in some cases.

Modular Assembly Versus Framework Formation

A biphasic transition (Figure 1 a) is evidence for modular assembly, or folding by parts. To use this as evidence for the mechanism of folding, it is necessary to know whether or not the molecule contains two or more stable domains. If so, the folding must be judged complex: the first goal is to understand the mechanism of folding for small "single-domain" proteins. Noncoincident transition curves (Figure 1 b), which show that the secondary structure is more stable than the tertiary structure, provide evidence for the framework model. Far-UV CD (210–240 nm) has been used as a probe of secondary structure, and either enzyme activity, specific ligand binding, or spectroscopic bands of aromatic residues (270–300 nm) can be used as probes of tertiary structure.

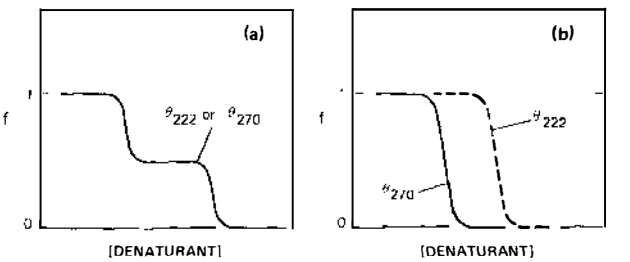

Figure 1 (*a*) Different probes of secondary and tertiary structure each give similar biphasic transition curves, which is interpreted as evidence for modular assembly or domain folding. (*b*) The secondary structure is more resistant than the tertiary structure to denaturant-induced unfolding, as measured by CD, which is interpreted as evidence for a framework model of folding.

α-Lactalbumin

The three-state ($N \rightleftharpoons I \rightleftharpoons U$) equilibrium unfolding transition of α-lactalbumin has been well characterized by Kuwajima, Sugai, and co-workers (101–108) and recently also by Ptitsyn and co-workers (224). The GuHCl unfolding transition shows non-coincident changes in CD at different wavelengths; aromatic signals (θ_{270} and θ_{296}) show an unfolding transition at a lower GuHCl concentration than the unfolding of the secondary structure (θ_{222}) (102, 104). Three important properties of I are: (*a*) I is in rapid equilibrium with U (a time constant of less than 1 msec), (*b*) I is present even when the protein's disulfide bonds are reduced, and (*c*) I has a far-UV CD spectrum close to that of N (102, 104). The fast interconversion ($U \rightleftharpoons I$) may be compared to the fast helix-coil transition of synthetic polypeptides (22–24). In contrast, the $I \rightleftharpoons N$ reaction can be measured in seconds (101, 103).

Similar intermediates (as judged by CD) are formed in 2 M GuHCl or 3 M $NaClO_4$ (105), and in unfolding by acid (102, 104) or base (108). Unfolding by acid or base involves the titration of ionizable groups that have abnormal pKs in N but normal pKs in I. These groups with abnormal pKs are responsible for a 10^4-fold increase in the unfolding rate ($N \rightarrow I$) at low and high pH (103, 108).

α-lactalbumin is not very stable (102). This may explain why an equilibrium intermediate is populated. The secondary structure may be intrinsically stable, while the tertiary structure is weak and easily disrupted by extremes of pH or by moderate concentrations of GuHCl or $NaClO_4$. It is interesting to compare α-lactalbumin with lysozyme, since these two proteins are believed to be structurally homologous (109). Lysozyme is a more stable protein, with a t_m approaching 80° at pH 5 (110). The high stability of its tertiary structure may prevent a comparable intermediate from being observed for lysozyme, because more drastic conditions are needed to unfold lysozyme than α-lactalbumin. Recently α-lactalbumin has been found to bind Ca^{2+} (107), and the stability of its tertiary structure is markedly increased by Ca^{2+} binding.

Penicillinase

Penicillinase is a single-chain protein (mol wt 29,000) with no disulfide bonds. Studies of the GuHCl-induced unfolding of penicillinase by Pain, Robson and co-workers (111–116) demonstrate at least one well-populated equilibrium intermediate and were initially interpreted as providing evidence for a framework model (111). The equilibrium transition curves fall into two categories (111, 112): (*a*) UV absorbance, aromatic CD, enzymatic activity, and viscosity measurements all show the same transition whereas

(*b*) far-UV CD and ORD changes occur at much higher GuHCl concentrations. The intermediate is a monomer, has significant secondary structure as judged by CD, and is expanded (112, 114). The existence of at least one equilibrium intermediate has been verified by NMR (116) and another intermediate has been found by urea-gradient electrophoresis (115). There are several similarities with results found for α-lactalbumin. (*a*) The intermediate has little tertiary structure (as measured with spectroscopic probes), but has a far-UV CD spectrum like that of N. (*b*) N is not very stable. (*c*) The intermediate is formed too rapidly to be measured with manual methods (i.e. within 30 sec), whereas the formation of N is slow (several minutes) (112, 116). Unlike penicillinase, the intermediate of α-lactalbumin is compact (224).

There is evidence for folding by domains (or by subdomains) in penicillinase. Digestion of the protein with CNBr results in three large fragments (mol wts 10,500, 9,500, and 8,500) that have been isolated. They combine to form a compact globular complex whose far-UV CD spectrum is indistinguishable from that of native penicillinase (113, 116). The reassembled complex has a different aromatic CD, and does not have enzymatic activity. The isolated fragments combine specifically with antibodies directed against native penicillinase, but it is not known whether they constitute stable domains by the test of showing a folding transition. The folding of penicillinase probably involves both framework formation and subdomain assembly.

Ovomucoid

Ovomucoid is a small (186-residue) protease inhibitor that is clearly a three-domain protein. The equilibrium unfolding transition is complex when measured by a single probe (117–121). Sequencing studies first indicated the presence of three homologous domains (121, 122). This explained earlier observations on the inhibitory properties of different avian ovomucoids; some have one site for trypsin inhibition, others have two sites, one each for trypsin and chymotrypsin; and still others have three sites, two for trypsin and one for chymotrypsin (123–125).

Domains can be isolated by digesting ovomucoid at interdomain sites (121, 126, 127). The isolated domains can refold independently into their native conformations. The isolated fragments: (*a*) have significant secondary and tertiary structure as judged by CD (128), (*b*) show both acid and thermal unfolding transitions (119, 121), (*c*) react with antisera to the native protein (126), (*d*) retain their inhibitory activity (121, 122), and (*e*) reform the correct disulfide bonds after reduction (120). One of the domains has recently been crystallized and the structure refined to 2.5 Å resolution (129). Recent analyses of the DNA and mRNA for ovomucoid

demonstrate that the gene segments coding for each domain are separated from each other by intervening sequences (130). The DNA sequence of the ovomucoid gene suggests that it evolved from a primordial ovomucoid gene by two separate intragenic duplications (130).

The unfolding transition of intact ovomucoid induced by GuHCl or urea is biphasic; both phases of the transition can be monitored by viscosity or tyrosine absorbance (117, 118). Since these probes usually monitor tertiary structure, it is not yet clear whether the two phases of the transition arise from independent unfolding of domains or from a two-step unfolding of the entire molecule (e.g. step 1: disruption of domain/domain contacts; step 2: unfolding of the individual domains).

Other Proteins

The GuHCl unfolding transition of growth hormone (mol wt 22,000) follows a framework model; changes in tertiary structure (A_{290}) occur at significantly lower GuHCl concentrations than changes in secondary structure (θ_{222}) (131). Moreover, as with α-lactalbumin, the intermediate is very similar to acid-denatured growth hormone (131, 132). Equilibrium studies on the GuHCl unfolding of carbonic anhydrase (mol wt of ~ 29,000) in the absence of Zn^{2+} also support a framework model (133), as do kinetic studies on this protein (78, 88). The equilibrium unfolding transition of cyt *c* shows spectrally measurable intermediates (134–139); loosening of the polypeptide chain around the heme precedes the unfolding of the remainder of the molecule. However, the calorimetric criterion for a two-state transition is satisfied with cyt *c* (99). The thermal unfolding transition of the *lac* repressor headpiece has been monitored by NMR; the changes in chemical shifts show that it is clearly not a two-state transition (140).

Proteins that have equilibrium properties suggesting domain assembly include: the α subunit of tryptophan synthase (141, 142), the β subunit of tryptophan synthase (143, 144), phosphorylase *b* (145), Bence Jones proteins (146), phosphoglycerate kinase (116), paramyosin (147), and myosin (148). In addition, papain, which has two structural domains separated by a deep cleft, gives a ratio of ΔH measured calorimetrically to the van't Hoff value ($\Delta H_{cal}/\Delta H_{vH}$) of 1.8, which suggests that the two domains fold almost independently (149).

Salt- and Methanol-Induced Unfolding

The role of neutral salts in stabilizing or denaturing folded proteins is complex (150–154). The effects of salts have been broken down into effects on the peptide group and effects on the nonpolar side chains (151, 153). Certain salts (e.g. LiCl, $LiClO_4$, and $CaCl_2$) induce unfolding of proteins at high concentrations. These unfolding transitions appear to be incomplete,

based on comparisons of the physical properties of the salt-denatured and GuHCl-denatured protein (16, 155, 156). Recently, it has been found that the addition of urea to salt-denatured RNase A produces a second transition (156).

Alcohols can also induce the unfolding of proteins (16). In particular, adding methanol or ethanol lowers the t_m of RNase A, and decreases the cooperativity of the transition (157, 158). Recently, proton NMR measurements at low temperatures in MeOH-H_2O mixtures have shown that equilibrium intermediates are well populated (158); the results have been interpreted by a framework model.

ENERGETICS OF FOLDING

These questions about the energetics of folding are of particular importance for understanding the mechanism of folding. (*a*) Which is the most stable structure of those proposed for the initial stage in folding? (*b*) Can the stabilities of possible folding intermediates be correlated with some property, such as water-accessible surface area, that can be computed directly from the structure (12, 41, 159)? (*c*) Are there specific structural interactions (e.g. H bonds or salt bridges) that are important energetically and that guide the formation of structure? Definitive answers will come from the structures of actual folding intermediates. Meanwhile, some specific questions are being answered from studies of model compounds and protein fragments. Also, accurate data are being obtained from studies of intact proteins on the factors that affect protein stability.

Model Compound Studies

In the last decade Scheraga and co-workers have determined the stability constant (s) and nucleation constant (σ) for α helix formation for most of the amino acid residues. A given residue is incorporated randomly as a "guest" in a water-soluble, helix-forming polypeptide "host" (derivatives of poly-L-glutamine). The results give the helix-forming propensities of the different amino acid residues as a function of temperature. A short, informative review of the method and results has been given (160).

The striking fact that emerges is that short α helices (10–20 residues) are intrinsically unstable in water. The ratio of helix to random coil is given approximately by: (helix)/(coil) $= \sigma s^{n-1}/(s-1)^2$ for short helices (175), where n is the number of amino acid residues. The largest value of s measured for any residue is 1.3 (Met at 0°C) (160). With $\sigma \leqslant 10^{-4}$, it is clear that a short helix of any composition should be unstable in water.

This conclusion had been predicted from studies of the stability of the amide (–NH...O–C–) H bond in aqueous solution (161, 162). Using an infrared technique, H-bonded dimers and higher oligomers could barely be

detected in aqueous solutions of N-methylacetamide (161). A lactam (δ-valerolactam), which forms two amide H bonds per dimer, forms a stronger dimer in water (163, 164), with an association constant at 25°C of 0.01 M^{-1} (164). This reaction of dimer formation shows a substantial, favorable enthalpy change: −3 kcal/mol H bond (164).[3]

A stronger H bond, which is easily measured in water by an NMR technique, is the charge stabilized H bond (–COO^-...HN–), formed when the side chain of a glutamic acid residue bends back to bond with its own peptide NH (165). The bond is disrupted by protonation of the carboxylate anion.

The free energy change for burying an amino acid side chain inside a protein as the protein folds up has been estimated from the free energy of transfer of the side chain from water to organic solvents (166). This transfer free energy has been correlated with the water-accessible surface area of hydrophobic side chains (12, 167, 168). Dispersion forces may make an important contribution to binding of hydrophobic amino acids to tRNA synthetases (169) and, similarly, they may be important in stabilizing protein folding. A theoretical study of the hydrophobic effect (170, 171) indicates that the correlation between reduction in water-accessible surface area and transfer free energy is not general and should not be extended to folding intermediates (171) without further justification. Transfer free energy data for the peptide group (151, 162) indicate that it strongly prefers to remain in water, and this should be considered in estimating the stability of a possible folding intermediate.

Studies with small molecules show that it is possible to demonstrate ion pairs in aqueous solution (e.g. guanidinium$^+$··· carboxylate$^-$ pairs) in a bimolecular reaction, but the strength of the interaction is not large (172, 173).

Protein Fragments

A decade ago several workers tested the possibility of using protein fragments as models for folding intermediates. If a protein folds first into microdomains, and these then coalesce into subdomains, and so on, then appropriately chosen fragments should fold at least partially, and their structures should give clues about the folding process. It is well documented

The free energy change for forming the peptide H bond in water has been estimated from the s and σ values for α helix formation (225). Since three residues are fixed in a helical conformation in the nucleation reaction without forming H bonds, we may take the free energy of the nucleation reaction (+ 5.6 kcal/mole if $\sigma = 10^{-4}$) and divide by 3 to get the free energy change per residue (+2 kcal/mole) when it adopts the helical conformation without forming an H bond. Since the s values obtained by the host-guest technique are close to 1, the overall free energy change including the H bond is close to 0, and the free energy change per H bond is −2 kcal/mol. This estimate does not distinguish between the entropic and enthalpic contributions to H bond formation.

that separate domains of larger proteins fold independently or nearly so [see below and also (5)]. However, most studies of subdomain fragments have given the disappointing result that the fragment is predominantly unfolded in aqueous solution.

Nevertheless, an intriguing example has been reported of a "microdomain" that shows partial helix formation in water (174, 175). The C peptide of RNase A (residues 1–13) is essentially unfolded in aqueous solution at 25°C but not at 1°C (174). Aggregation occurs above 2 mg/ml but the helix forms intramolecularly, and helix formation is observed by CD at concentrations as low as 40 μg/ml. NMR data indicate that all, or nearly all, residues participate in helix formation (175). Up to $\sim$ 30% helix content can be observed in water (175), which is 1000 times greater than the helix content predicted by the host-guest studies. pH titration shows that protonation of His 12, and deprotonation of either Glu 2 or Glu 9 or of both are required for significant helix formation (175). The results suggest that specific salt bridge(s) (e.g. His 12^+...Glu 9^-) nucleate the helix by stabilizing the first turn.

Intact Proteins

Salt bridges have been suspected of being important for folding since their discovery in α-chymotrypsin (176) and hemoglobin (177–179). Recently, salt bridges have been demonstrated in several proteins, including phosphorylase (180) and BPTI (181). Estimates of the strength of individual salt bridges range from –1 to –3 kcal/mol (179, 181, 182).

An advance has been made in treating the electrostatic properties of proteins. A simple discrete charge model, which makes use of X-ray structures to give the locations and solvent accessibilities of the ionizing groups, predicts rather well the pKs of individual groups observed by NMR [(183, 184), see however (185)]. In doing this the model also predicts the electrostatic contribution to the free energy of folding.

The overall thermodynamics of folding are now known accurately from calorimetry for several proteins (10). The data confirm that the folded structures of globular proteins are only marginally stable ($\Delta G = -5$ to -15 kcal/mol), and demonstrate that the thermodynamics of folding as a function of temperature are dominated by the large and approximately constant value of ΔC_P: both ΔH and ΔS are strongly temperature dependent (10, 186).

The striking conclusion that emerges from these studies is that there is a large and favorable contribution to the enthalpy of unfolding that cannot come from hydrophobic interactions (10). Privalov argues that this nonhydrophobic contribution to the unfolding enthalpy is nearly temperature independent and that it arises from H bonds and dispersion forces (10). If one assumes that it arises entirely from H bond enthalpy, then its magnitude

corresponds to about –1.7 kcal/mol H bond for several proteins (10), which can be compared with the –3 kcal/mol H bond estimated for the dimerization of δ-valerolactam (163, 164). These results suggest that peptide H bonds may make a major contribution to the stabilization of the native structure.

RELATED TOPICS

The following topics are closely related to our review but limitations of space prevent their review here. We give references to reviews and to some recent papers, and comment on the relationships to work reviewed here.

Disulfide Intermediates

The best evidence for specific intermediates in protein folding experiments comes from the studies by Creighton of trapped disulfide intermediates in BPTI. The work has been reviewed recently (3, 187). Some basic properties of the system are as follows. (*a*) There are multiple intermediates and multiple pathways: however, there is a "most-favored" pathway and obligatory intermediates. (*b*) The obligatory two-disulfide intermediates each have one non-native S–S bond. Moreover, an abortive, or dead-end, intermediate has two native S–S bonds. Theorists working with computer simulation of folding have sought to explain these surprising facts. (*c*) The spectrum of intermediates narrows down toward the most-favored pathway in conditions favoring folding (188) (i.e. low temperatures, or the presence of stabilizing Hofmeister anions such as SO_4^{2-}). This increases the overall rate of the refolding/reoxidation process. Thus, the most-favored pathway proceeds via the most stable intermediates, and the rate of the overall process depends on how well these intermediates are populated. (*d*) The one-disulfide intermediates equilibrate with each other before the second S–S bond is formed and so do the two-disulfide intermediates, albeit more slowly [compare the recent study of RNase A by Konishi et al (189)]. To a first approximation, there is "thermodynamic control" of the refolding/reoxidation pathways. (*e*) The major folding process occurs in a single S–S rearrangement, near the end of the pathway. The study of earlier trapped intermediates has shown chiefly that it is difficult to detect and characterize any specific structure (190). Nevertheless, the importance of specific interactions is shown by the close correlation between a narrow or broad spectrum of these intermediates and whether the folding conditions are strongly native or marginally so (188). (*f*) Unfolding/reduction and refolding/reoxidation can be studied in the same conditions by varying the ratio of oxidant to reductant. Since the conditions are the same and the process is reversible, the pathways of unfolding and refolding are the same.

Domain Folding and Exons

Stable domains that show folding transitions have been isolated after limited proteolysis from numerous large proteins, and the roles of these domains in folding have been investigated in a few cases, including the β_2 subunit of tryptophan synthase (143, 144, 191), the "double-headed" enzyme aspartokinase-homoserine dehydrogenase I of *E. coli,* which has both enzyme activities joined in a single chain (192–194), immunoglobulins, including Bence-Jones proteins (146, 195, 196), elastase (197), and the λ repressor (198). A generalization from these studies is that domains often fold independently in kinetic terms, and are thermodynamically stable, but that a subsequent slow rearrangement, which involves interactions between domains, is commonly required for full activity. We have already discussed domain folding in the case of the small proteins ovomucoid, penicillinase, and the α subunit of tryptophan synthase.

Gilbert (199, 200) has proposed that exons code for protein domains and that genetic recombination within introns speeds up the evolution of new proteins. Introns that separate domain coding regions have been demonstrated for immunoglobulins (201, 202) and for ovomucoid (130). A large intron occurs inside the coding region for the C peptide, thus separating the A and B chains of rat insulin (203). Introns occur within the coding regions for α helices in the globin genes (204): however, the polypeptide fragment coded by the central exon does bind heme specifically and tightly (205).

Oligomeric Proteins

A systematic study of the folding and assembly of several oligomeric enzymes has been made by Jaenicke and co-workers; this work has been reviewed recently (206). Some central points are as follows. (*a*) Only in rare cases is thermodynamic equilibrium ever reached between unfolded monomer and folded oligomer. Kinetic studies give information about a folding pathway that is not readily reversible in most cases. Nevertheless, refolding can give native enzyme in almost quantitative yield in special conditions. (*b*) Aggregation of partly folded chains is the major technical problem. It can be minimized by special procedures, including folding conditions that stabilize folded monomers. (*c*) In general, folded monomers are inactive; rabbit muscle aldolase is an exception. (*d*) Specificity of association is high: mixtures of closely related enzymes (e.g. lactate and malate dehydrogenase) do not refold to give "chimeric" species (207).

The pathways of assembly of aspartate transcarbamylase from catalytic and regulatory subunits have been studied by Schachman & co-workers (208–210). Pulse-chase experiments with radioactively labeled subunits, followed by separation of intermediates in electrophoresis, have been used to give a model for the assembly process.

Temperature-sensitive mutants in the assembly or folding of the phage P22 tail spike protein have been studied by King and co-workers (211–213). These mutants appear to be kinetically blocked in folding or assembly at restrictive temperatures. When synthesized at the permissive temperature, the mutants are as stable as the wild-type protein, as measured by the kinetics of irreversible thermal denaturation. When synthesized at the restrictive temperature, many of the mutant proteins can be reactivated in the absence of new protein synthesis by a shift to the permissive temperature, which indicates that the mutant chains that accumulate at the restrictive temperature are capable of folding and assembly (213).

Fragment Exchange and Local Unfolding Reactions

Certain pairs of polypeptide fragments, which individually are unfolded, can combine to form a native-like, enzymatically active, complex. This type of complementation has been studied extensively by Taniuchi and co-workers for two small proteins, staph nuclease (20) and cyt *c* (214). The probability of successful complementation increases if the fragments are overlapping. Even three fragments can combine to form a complex (215). The dissociation of these complexes is of particular interest as a model system for studying local unfolding reactions. Local unfolding has often been proposed as a chief means of allowing amide proton exchange in native proteins (216), and local unfolding reactions could be part of the overall unfolding pathway.

Current views of protein flexibility indicate that water can penetrate readily into the interior of a globular protein, and it has been suggested that minor perturbations of the folded structure may permit amide proton exchange (217–219). To decide between "deep breathing" and highly local breathing as dominant mechanisms of exchange, it is necessary to measure the kinetics and equilibria of the breathing reactions by independent methods. Fragment exchange studies can yield both kinetic and equilibrium data and, since the contact regions between fragments are extensive, it is fairly certain that deep breathing reactions are required to break the contacts. A major conclusion from the fragment exchange studies is that the rates and equilibrium constants are in a range where they can be expected to contribute significantly to amide proton exchange in native proteins. This has been demonstrated directly for the dissociation of S peptide from RNase S (220, 221) where amide proton exchange of ^{3}H-labeled S peptide combined with S protein is concentration dependent, and therefore occurs partly by dissociation to free S peptide, even at 0°C, pH 7. A local unfolding reaction in staph nuclease has been demonstrated directly by using antibodies directed against the unfolded protein (222).

CONCLUDING REMARKS

The kinetic mechanism of protein folding proves to be *sequential folding* with defined intermediates in two cases: the major folding reaction of RNase A ($U_S^{II} \rightarrow N$) and the refolding/reoxidation of reduced BPTI or of reduced RNase A. It is likely that sequential folding is a general mechanism. The main objection to it has been the high cooperativity of folding measured by equilibrium experiments inside the folding transition zone, which implies that all folding intermediates are unstable. This objection does not apply if the kinetics of folding are studied outside the transition zone, where intermediates can be stable relative to the unfolded protein. Finding small proteins that do show equilibrium intermediates (especially α-lactalbumin, which is probably a single-domain protein) adds to the evidence that the cooperativity of folding is marginal, not absolute.

Understanding the role of *proline isomerization* in the kinetics of folding has been an essential part of the recent progress. In order to study the kinetic pathway of folding, it is necessary to identify the different unfolded species of a protein and to study the folding of each one separately.

Most present results support a *framework model* of folding in which the H-bonded secondary structure is formed at an early stage, and they are not consistent with strict *modular assembly* of small proteins, or folding by parts, in which both the secondary and tertiary structures of any part of a protein are formed at the same time. However, the H-bonded secondary structure even of a small protein may itself be formed in distinct stages. Modular assembly may apply to the domain folding of larger, multidomain proteins; however, present evidence suggests that a structural rearrangement occurs after the initial folding of separate domains and before the full activities of the native protein are regained. The framework model was first suggested by finding that the secondary structures of some unusual small proteins (e.g. penicillinase, carbonic anhydrase, and α-lactalbumin) are more resistant to unfolding by denaturants than are their tertiary structures. Kinetic intermediates consistent with a framework model were found by ^3H trapping experiments with RNase A, which showed that many NH protons are protected from exchange with solvent early in folding and then throughout the folding process, as expected if H-bonded secondary structure is formed early in folding. Stopped-flow CD measurements appear capable of resolving stages in the formation of α helices and β sheets, during the $U_S \rightarrow N$ folding reaction of RNase S.

Well-populated intermediates in the folding of small proteins are now a reality. Characterization of these intermediates will provide a benchmark for theorists working on prediction of folding from sequence and the elucidation of the folding pathway.

ACKNOWLEDGMENTS

We thank Drs. K. Kuwajima and O. B. Ptitsyn for discussion of this review, and many colleagues for sending manuscripts not yet in print.

P. S. Kim is a predoctoral fellow of the Medical Scientist Training Program supported by the National Institutes of Health (GM 07365). This work has been supported by grants to R. L. Baldwin from the National Science Foundation (PCM 77-16834) and the National Institutes of Health (2 RO1 GM 19988–21). Use of the Stanford Magnetic Resonance Facility (supported by NSF Grant GP 23633 and NIH Grant RR 00711) is gratefully acknowledged.

Literature Cited

1. Jaenicke, R., ed. 1980. *Protein Folding, Proc. 28th Conf. German Biochem. Soc.* Amsterdam: Elsevier/North-Holland Biomed. Press. 587 pp.
1a. Ptitsyn, O. B. 1981. *FEBS Lett.* 131: 197–202
2. Ghélis, C., Yon, J. 1982. *Protein Folding.* New York: Academic. In press
3. Creighton, T. E. 1978. *Prog. Biophys. Mol. Biol.* 33:231–97
4. Baldwin, R. L. 1975. *Ann. Rev. Biochem.* 44:453–75
5. Wetlaufer, D. B. 1981. *Adv. Protein Chem.* 34:61–92
6. Richardson, J. S. 1981. *Adv. Protein Chem.* 34:167–339
7. Ptitsyn, O. B., Finkelstein, A. V. 1980. *Q. Rev. Biophys.* 13:339–86
8. Thomas, K. A., Schechter, A. N. 1980. In *Biological Regulation and Development,* ed. R. F. Goldberger, 2:43–100. New York: Plenum
9. Rossmann, M. G., Argos, P. 1981. *Ann. Rev. Biochem.* 50:497–532
10. Privalov, P. L. 1979. *Adv. Protein Chem.* 33:167–241
11. Némethy, G., Scheraga, H. A. 1977. *Q. Rev. Biophys.* 10:239–352
12. Richards, F. M. 1977. *Ann. Rev. Biophys. Bioeng.* 6:151–76
13. Ikegami, A. 1981. *Adv. Chem. Phys.* 46:363–413
14. Anfinsen, C. B., Scheraga, H. A. 1975. *Adv. Protein Chem.* 29:205–300
15. Wetlaufer, D. B., Ristow, S. 1973. *Ann. Rev. Biochem.* 42:135–58
16. Tanford, C. 1968. *Adv. Protein Chem.* 23:121–282
17. Tanford, C. 1970. *Adv. Protein Chem.* 24:1–95
18. Levinthal, C. 1968. *J. Chim. Phys.* 65: 44–45
19. Karplus, M., Weaver, D. L. 1976. *Nature* 260:404–6
20. Taniuchi, H. 1973. *J. Biol. Chem.* 248: 5164–74
21. Levitt, M., Warshel, A. 1975. *Nature* 253:694–98
22. Hammes, G. G., Roberts, P. B. 1969. *J. Am. Chem. Soc.* 91:1812–16
23. Barksdale, A. D., Stuehr, J. E. 1972. *J. Am. Chem. Soc.* 94:3334–38
24. Gruenewald, B., Nicola, C. U., Lustig, A., Schwarz, G., Klump, H. 1979. *Biophys. Chem.* 9:137–47
25. Ptitsyn, O. B., Rashin, A. A. 1975. *Biophys. Chem.* 3:1–20
26. Karplus, M., Weaver, D. L. 1979. *Biopolymers* 18:1421–37
27. Cohen, F. E., Sternberg, M. J. E., Phillips, D. C., Kuntz, I. D., Kollman, P. A. 1980. *Nature* 286:632–34
28. Hammes, G. G. 1978. *Principles of Chemical Kinetics.* New York: Academic. 268 pp.
29. Tsong, T. Y., Baldwin, R. L. 1978. *Biopolymers* 17:1669–78
30. Baldwin, R. L. 1980. See Ref. 1, pp. 369–85
31. Haas, E., Katchalski-Katzir, E., Steinberg, I. Z. 1978. *Biopolymers* 17:11–31
32. Tsong, T. Y. 1982. *Biochemistry.* In press
33. Cook, K. H., Schmid, F. X., Baldwin, R. L. 1979. *Proc. Natl. Acad. Sci. USA* 76:6157–61
34. Schmid, F. X., Baldwin, R. L. 1979. *J. Mol. Biol.* 135:199–215
35. Kim, P. S., Baldwin, R. L. 1980. *Biochemistry* 19:6124–29
36. Schmid, F. X. 1980. See Ref. 1, pp. 387–400
37. Schmid, F. X. 1981. *Eur. J. Biochem.* 114:105–9

38. Schmid, F. X., Blaschek, H. 1981. *Eur. J. Biochem.* 114:111–17
39. Labhardt, A. M., Baldwin, R. L. 1979. *J. Mol. Biol.* 135:231–44
40. Labhardt, A. M. 1980. See Ref. 1, pp. 401–25
41. Rashin, A. A. 1979. *Stud. Biophys.* 77:177–84
42. Wako, H., Saitô, H. 1978. *J. Phys. Soc. Jpn.* 44:1931–38
43. Abe, H., Gō, N. 1981. *Biopolymers* 20: 1013–31
44. Gō, N., Abe, H. 1981. *Biopolymers* 20: 991–1011
45. Lim, V. 1980. See Ref. 1, pp. 149–66
46. Lesk, A. M., Rose, G. D. 1981. *Proc. Natl. Acad. Sci. USA* 78:4304–8
47. Richardson, J. S. 1977. *Nature* 268: 495–500
48. Brandts, J. F., Halvorson, H. R., Brennan, M. 1975. *Biochemistry* 14:4953–63
49. Garel, J.-R., Baldwin, R. L. 1973. *Proc. Natl. Acad. Sci. USA* 70:3347–51
50. Nall, B. T., Garel, J.-R., Baldwin, R. L. 1978. *J. Mol. Biol.* 118:317–30
51. Schmid, F. X., Baldwin, R. L. 1978. *Proc. Natl. Acad. Sci. USA* 75:4764–68
52. Hagerman,P. J., Schmid, F. X., Baldwin, R. L. 1979. *Biochemistry* 18: 293–97
53. Hagerman, P. J., Baldwin, R. L. 1976. *Biochemistry* 15:1462–73
54. Garel, J.-R., Nall, B. T., Baldwin, R. L. 1976. *Proc. Natl. Acad. Sci. USA* 73: 1853–57
55. Kato, S., Okamura, M., Shimamoto, N., Utiyama, H. 1981. *Biochemistry* 20:1080–85
56. Kato, S., Shimamoto, N., Utiyama, H. 1982. *Biochemistry.* 21:38–43
57. Ridge, J. A., Baldwin, R. L., Labhardt, A. M. 1981. *Biochemistry* 20:1622–30
58. Creighton, T. E. 1980. *J. Mol. Biol.* 137:61–80
59. Jullien, M., Baldwin, R. L. 1981. *J. Mol. Biol.* 145:265–80
60. Nall, B. T., Landers, T. A. 1981. *Biochemistry* 20:5403–11
61. McPhie, P. 1982. *J. Biol. Chem.* 257:689–93
62. Grathwohl, C., Wüthrich, K. 1976. *Biopolymers* 15:2025–41
63. Grathwohl, C., Wüthrich, K. 1976. *Biopolymers* 15:2043–57
64. Cheng, H. N., Bovey, F. A. 1977. *Biopolymers* 16:1465–72
65. Steinberg, I. Z., Harrington, W. F., Berger, A., Sela, M., Katchalski, E. 1960. *J. Am. Chem. Soc.* 82:5263–79
66. Garel, J.-R. 1980. *Proc. Natl. Acad. Sci. USA* 77:795–98
67. Grathwohl, C., Wüthrich, K. 1981. *Biopolymers* 20:2623–33
68. Brandts, J. F., Brennan, M., Lin, L.-N. 1977. *Proc. Natl. Acad. Sci. USA* 74: 4178–81
69. Lin, L.-N., Brandts, J. F. 1978. *Biochemistry* 17:4102–10
70. Babul, J., Nakagawa, A., Stellwagen, E. 1978. *J. Mol. Biol.* 126:117–21
71. Bruckner, P., Bächinger, H. P., Timpl, R., Engel, J. 1978. *Eur. J. Biochem. 90,* 595–604
72. Levitt, M. 1981. *J. Mol. Biol.* 145: 251–63
73. Wüthrich, K., Grathwohl, C. 1974. *FEBS Lett.* 43:337–40
74. Hetzel, R., Wüthrich, K. 1979. *Biopolymers* 18:2589–606
75. Rees, D. C., Lewis, M., Honzatko, R. B., Lipscomb, W. N., Hardman, K. D. 1981. *Proc. Natl. Acad. Sci. USA* 78: 3408–12
76. Luchins, J., Beychok, S. 1978. *Science* 199:425–26
77. Kihara, H., Takahashi, E., Yamamura, K., Tabushi, I. 1980. *Biochem. Biophys. Res. Commun.* 95:1687–94
78. Ko, B. P. N., Yazgan, A., Yeagle, P. L., Lottich, S. C., Henkens, R. W. 1977. *Biochemistry* 16:1720–25
79. Willumsen, L. 1971. *CR Trav. Lab. Carlsberg* 38:223–95
80. Karplus, S., Snyder, G. H., Sykes, B. D. 1973. *Biochemistry* 12:1323–29
81. Rehage, A., Schmid, F. X. 1982. *Biochemistry.* In press
82. Labhardt, A. M., Baldwin, R. L. 1979. *J. Mol. Biol.* 135:245–54
83. Tanford, C., Aune, K. C., Ikai, A. 1973. *J. Mol. Biol.* 73:185–97
84. Hagerman, P. J. 1977. *Biopolymers* 16:731–47
85. Ikai, A., Fish, W. F., Tanford, C. 1973. *J. Mol. Biol.* 73:165–84
86. Tsong, T. Y. 1973. *Biochemistry* 12: 2209–14
87. Stein, P. J., Henkens, R. W. 1978. *J. Biol. Chem.* 253:8016–18
88. McCoy, L. F., Rowe, E. S., Wong, K.-P. 1980. *Biochemistry* 19:4738–43
89. Carlsson, U., Aasa, R., Henderson, L. E., Jonsson, B.-H., Lindskog, S. 1975. *Eur. J. Biochem.* 52:25–36
90. Yazgan, A., Henkens, R. W. 1972. *Biochemistry* 11:1314–18
91. Wong, K.-P., Hamlin, L. M. 1975. *Arch. Biochem. Biophys.* 170:12–22
92. Nilsson, A., Lindskog, S. 1967. *Eur. J. Biochem.* 2:309–17
93. Wong, K.-P., Hamlin, L. M. 1974. *Biochemistry* 13:2678–83

94. Ikai, A., Tanaka, S., Noda, H. 1978. *Arch. Biochem. Biophys.* 190:39–45
95. Matthews, C. R., Crisanti, M. M. 1981. *Biochemistry* 20:784–92
96. Crisanti, M. M., Matthews, C. R. 1981. *Biochemistry* 20:2700–6
97. Higgins, W., Fairwell, T., Miles, E. W. 1979. *Biochemistry* 18:4827–35
98. Leutzinger, Y., Beychok, S. 1981. *Proc. Natl. Acad. Sci. USA* 78:780–84
99. Privalov, P. L., Khechinashvili, N. N. 1974. *J. Mol. Biol.* 86:665–84
100. Privalov, P. L. 1974. *FEBS Lett.* 40: (Suppl) S140–53
101. Kuwajima, K., Nitta, K., Sugai, S. 1975. *J. Biochem. Tokyo* 78:205–11
102. Kuwajima, K., Nitta, K., Yoneyama, M., Sugai, S. 1976. *J. Mol. Biol.* 106: 359–73
103. Kita, N., Kuwajima, K., Nitta, K., Sugai, S. 1976. *Biochim. Biophys. Acta* 427:350–58
104. Kuwajima, K. 1977. *J. Mol. Biol.* 114: 241–58
105. Maruyama, S., Kuwajima, K., Nitta, K., Sugai, S. 1977. *Biochem. Biophys. Acta* 494:343–53
106. Nozaka, M., Kuwajima, K., Nitta, K., Sugai, S. 1978. *Biochemistry* 17: 3753–58
107. Hiraoka, Y., Segawa, T., Kuwajima, K., Sugai, S., Murai, N. 1980. *Biochem. Biophys. Res. Commun.* 95:1098–1104
108. Kuwajima, K., Ogawa, Y., Sugai, S. 1981. *J. Biochem. Tokyo* 89:759–70
109. Browne, W. J., North, A. C. T., Phillips, D. C., Brew, K., Vanaman, T. C., Hill, R. L. 1969. *J. Mol. Biol.* 42:65–86
110. Khechinashvili, N. N., Privalov, P. L., Tiktopulo, E. I. 1973. *FEBS Lett.* 30:57–60
111. Robson, B., Pain, R. H. 1973. In *Conformation of Biological Molecules Polymers, 5th Jerusalem Symp.* ed. E. D. Bergmann, B. Pullman, pp. 161–72. New York: Academic
112. Robson, B., Pain, R. H. 1976. *Biochem. J.* 155:331–44
113. Carrey, E. A., Pain, R. H. 1977. *Biochem. Soc. Trans.* 5:689–92
114. Carrey, E. A., Pain, R. H. 1978. *Biochim. Biophys. Acta* 533:12–22
115. Creighton, T. E., Pain, R. H. 1980. *J. Mol. Biol.* 137:431–36
116. Adams, B., Burgess, R. J., Carrey, E. A., Mackintosh, I. R., Mitchinson, C., Thomas, R. M., Pain, R. H. 1980. See Ref. 1, pp. 447–67
117. Waheed, A., Qasim, M. A., Salahuddin, A. 1977. *Eur. J. Biochem.* 76:383–90
118. Baig, M. A., Salahuddin, A. 1978. *Biochem. J.* 171:89–97
119. Matsuda, T., Watanabe, K., Sato, Y. 1981. *Biochem. Biophys. Acta* 669: 109–12
120. Matsuda, T., Watanabe, K., Sato, Y. 1981. *FEBS Lett.* 124:185–88
121. Kato, I., Kohr, W. J., Laskowski, M. 1978. *FEBS Meet.* 47:197–206
122. Kato, I., Schrode, J., Wilson, K. A., Laskowski, M. 1976. In *Protides of the Biological Fluids, 23rd,* ed. H. Peeters, pp. 235–43. Oxford: Pergamon. 699 pp.
123. Rhodes, M. B., Bennett, N., Feeney, R. E. 1960. *J. Biol. Chem.* 235:1686–93
124. Osuga, D. T., Bigler, J. C., Uy, R. L., Sjøberg, L., Feeney, R. E. 1974. *Comp. Biochem. Physiol. B* 48:519–33
125. Laskowski, M., Kato, I. 1980. *Ann. Rev. Biochem.* 49:593–626
126. Beeley, J. G. 1976. *Biochem. J.* 155:345–51
127. Bogard, W. C., Kato, I., Laskowski, M. 1980. *J. Biol. Chem.* 255:6569–74
128. Watanabe, K., Matsuda, T., Sato, Y. 1981. *Biochem. Biophys. Acta* 667: 242–50
129. Weber, E., Papamokos, E., Bode, W., Huber, R., Kato, I., Laskowski, M. 1981. *J. Mol. Biol.* 149:109–23
130. Stein, J. P., Catterall, J. F., Kristo, P., Means, A. R., O'Malley, B. W. 1980. *Cell* 21:681–87
131. Holladay, L. A., Hammonds, R. G., Puett, D. 1974. *Biochemistry* 13: 1653–61
132. Burger, H. G., Edelhoch, H., Condliffe, P. G. 1966. *J. Biol. Chem.* 241:449–57
133. Wong, K.-P., Tanford, C. 1973. *J. Biol. Chem.* 248:8518–23
134. Schejter, A., George, P. 1964. *Biochemistry* 3:1045–49
135. Myer, Y. P. 1968. *Biochemistry* 7: 765–76
136. Stellwagen, E. 1968. *Biochemistry* 7: 2893–98
137. Kaminsky, L. S., Miller, V. J., Davison, A. J. 1973. *Biochemistry* 12:2215–21
138. Drew, H. R., Dickerson, R. E. 1978. *J. Biol. Chem.* 253:8420–27
139. Myer, Y. P., MacDonald, L. H., Verma, B. C., Pande, A. 1980. *Biochemistry* 19:199–207
140. Wemmer, D., Ribeiro, A. A., Bray, R. P., Wade-Jardetzky, N. G., Jardetzky, O. 1981. *Biochemistry* 20:829–33
141. Yutani, K., Ogasahara, K., Suzuki, M., Sugino, Y. 1979. *J. Biochem. Tokyo* 85:915–21
142. Yutani, K., Ogasahara, K., Sugino, Y. 1980. *J. Mol. Biol.* 144:455–65
143. Goldberg, M. E., Zetina, C. R. 1980. See Ref. 1, pp. 469–84

144. Zetina, C. R., Goldberg, M. E. 1980. *J. Mol. Biol.* 137:401–14
145. Chignell, D. A., Azhir, A., Gratzer, W. B. 1972. *Eur. J. Biochem.* 26:37–42
146. Azuma, T., Hamaguchi, K., Migita, S. 1972. *J. Biochem. Tokyo* 72:1457–67
147. Riddiford, L. M. 1966. *J. Biol. Chem.* 241:2792–802
148. Tsong, T. Y., Karr, T., Harrington, W. F. 1979. *Proc. Natl. Acad. Sci. USA* 76:1109–13
149. Tiktopulo, E. I., Privalov, P. L. 1978. *FEBS Lett.* 91:57–58
150. von Hippel, P. H., Wong, K. Y. 1965. *J. Biol. Chem.* 240:3909–23
151. Nandi, P. K., Robinson, D. R. 1972. *J. Am. Chem. Soc.* 94:1299–1315
152. Schrier, E. E., Schrier, E. B. 1967. *J. Phys. Chem.* 71:1851–60
153. von Hippel, P. H., Peticolas, V., Schack, L., Karlson, L. 1973. *Biochemistry* 12:1256–64
154. Melander, W., Horváth, C. 1977. *Arch. Biochem. Biophys.* 183:200–15
155. Sharma, R. N., Bigelow, C. C. 1974. *J. Mol. Biol.* 88:247–57
156. Ahmad, F., Bigelow, C. C. 1979. *J. Mol. Biol.* 131:607–17
157. Fink, A. L., Grey, B. L. 1978. In *Biomolecular Structure and Function* eds. P. F. Agris, R. N. Loeppky, B. D. Sykes, pp. 471–77. New York: Academic. 614 pp.
158. Biringer, R. G., Fink, A. L. 1982. *J. Mol. Biol.* In press
159. Lesk, A. M., Chothia, C. 1980. In *Biophysical Discussions, Proteins and Nucleoproteins, Structure, Dynamics, and Assembly,* ed. V. A. Parsegian. pp. 35–47. New York: Rockefeller Univ. Press. 676 pp.
160. Scheraga, H. A. 1978. *Pure Appl. Chem.* 50:315–24
161. Klotz, I. M., Franzen, J. S. 1962. *J. Am. Chem. Soc.* 84:3461–66
162. Klotz, I. M., Farnham, S. B. 1968. *Biochemistry* 7:3879–82
163. Susi, H., Timasheff, S. N., Ard, J. S. 1964. *J. Biol. Chem.* 239:3051–54
164. Susi, H. 1969. In *Structure and Stability of Biological Macromolecules,* eds. S. N. Timasheff, G. D. Fasman, pp. 2:575–663. New York: Dekker. 694 pp.
165. Bundi, A., Wüthrich, K. 1979. *Biopolymers* 18:299–311
166. Nozaki, Y., Tanford, C. 1971. *J. Biol. Chem.* 246:2211–17
167. Chothia, C. 1975. *Nature* 254:304–8
168. Reynolds, J. A., Gilbert, D. B., Tanford, C. 1974. *Proc. Natl. Acad. Sci. USA* 71:2925–27
169. Fersht, A. R., Shindler, J. S., Tsui, W.-C. 1980. *Biochemistry* 19:5520–24
170. Pratt, L. R., Chandler, D. 1977. *J. Chem. Phys.* 67:3683–704
171. Karplus, M. 1980. See Ref. 159, pp. 45–46
172. Tanford, C. 1954. *J. Am. Chem. Soc.* 76:945–46
173. Springs, B., Haake, P. 1977. *Bioorg. Chem.* 6:181–90
174. Brown, J. E., Klee, W. A. 1971. *Biochemistry* 10:470–76
175. Bierzynski, A., Kim, P. S., Baldwin, R. L. 1982. *Proc. Natl. Acad. Sci. USA.* In press
176. Birktoft, J. J., Blow, D. M. 1972. *J. Mol. Biol.* 68:187–240
177. Perutz, M. F., TenEyck, L. F. 1972. *Cold Spring Harbor Symp. Quant. Biol.* 36:295–310
178. Kilmartin, J. V., Hewitt, J. A. 1972. *Cold Spring Harbor Symp. Quant. Biol.* 36:311–14
179. Perutz, M. F. 1978. *Science* 201: 1187–91
180. Sprang, S., Fletterick, R. J. 1980. See Ref. 159, pp. 175–92
181. Brown, L. R., DeMarco, A., Richarz, R., Wagner, G., Wüthrich, K. 1978. *Eur. J. Biochem.* 88:87–95
182. Fersht, A. R. 1972. *Cold Spring Harbor Symp. Quant. Biol.* 36:71–73
183. Shire, S. J., Hanania, G. I. H., Gurd, F. R. N. 1975. *Biochemistry* 14:1352–58
184. Matthew, J. B., Hanania, G. I. H., Gurd, F. R. N. 1979. *Biochemistry* 18:1919–36
185. Kilmartin, J. V., Fogg, J. H., Perutz, M. F. 1980. *Biochemistry* 19:3189–93
186. Schellman, J. A., Hawkes, R. B. 1980. See Ref. 1, pp. 331–43
187. Creighton, T. E. 1980. See Ref. 1, pp. 427–46
188. Creighton, T. E. 1980. *J. Mol. Biol.* 144:521–50
189. Konishi, Y., Ooi, T., Scheraga, H. A. 1981. *Biochemistry* 20:3945–55
190. Kosen, P. A., Creighton, T. E., Blout, E. R. 1980. *Biochemistry* 19:4936–44
191. Högberg-Raibaud, A., Goldberg, M. E. 1977. *Proc. Natl. Acad. Sci. USA* 74:442–46
192. Garel, J.-R., Dautry-Varsat, A. 1980. See Ref. 1, pp. 485–99
193. Garel, J.-R., Dautry-Varsat, A. 1980. *Proc. Natl. Acad. Sci. USA* 77:3379–83
194. Dautry-Varsat, A., Garel, J.-R. 1981. *Biochemistry* 20:1396–401
195. Isenman, D. E., Lancet, D., Pecht, I. 1979. *Biochemistry* 18:3327–36
196. Goto, Y., Hamaguchi, K. 1981. *J. Mol. Biol.* 146:321–40

197. Ghélis, C., Tempete-Gaillourdet, M., Yon, J. M. 1978. *Biochem. Biophys. Res. Commun.* 84:31–36
198. Pabo, C. O., Sauer, R. T., Sturtevant, J. M., Ptashne, M. 1979. *Proc. Natl. Acad. Sci. USA* 76:1608–12
199. Gilbert, W. 1978. *Nature* 271:501
200. Gilbert, W. 1980. *J. Supramolec. Structure* Suppl. 4, p. 115. (Abstr. 9th ICN-UCLA Symp.)
201. Brack, C., Tonegawa, S. 1977. *Proc. Natl. Acad. Sci. USA* 74:5652–56
202. Brack, C., Hirama, M., Lenhard-Schuller, R., Tonegawa, S. 1978. *Cell* 15:1–14
203. Lomedico, P., Rosenthal, N., Efstratiadis, A., Gilbert, W., Kolodner, R., Tizard, R. 1979. *Cell* 18:545–58
204. Lawn, R. M., Fritsch, E. F., Parker, R. C., Blake, G., Maniatis, T. 1978. *Cell* 15:1157–74
205. Craik, C. S., Buchman, S. R., Beychok, S. 1980. *Proc. Natl. Acad. Sci. USA* 77:1384–88
206. Jaenicke, R., Rudolph, R. 1980. See Ref. 1, pp. 525–48
207. Jaenicke, R., Rudolph, R., Heider, I. 1981. *Biochem. Int.* 2:23–31
208. Bothwell, M., Schachman, H. K. 1974. *Proc. Natl. Acad. Sci. USA* 71:3221–25
209. Bothwell, M. A., Schachman, H. K. 1980. *J. Biol. Chem.* 255:1962–70
210. Bothwell, M. A., Schachman, H. K. 1980. *J. Biol. Chem.* 255:1971–77
211. Smith, D. H., Berget, P. B., King, J. 1980. *Genetics* 96:331–52
212. Goldenberg, D. P., King, J. 1981. *J. Mol. Biol.* 145:633–51
213. Smith, D. H., King, J. 1981. *J.Mol. Biol.* 145:653–76
214. Hantgan, R. R., Taniuchi, H. 1978. *J. Biol. Chem.* 253:5373–80
215. Andria, G., Taniuchi, H., Cone, J. L. 1971. *J. Biol. Chem.* 246:7421–28
216. Englander, S. W., Downer, N. W., Teitelbaum, H. 1972. *Ann. Rev. Biochem.* 41:903–24
217. Woodward, C. K., Hilton, B. D. 1979. *Ann. Rev. Biophys. Bioeng.* 8:99–127
218. Richards, F. M. 1979. *Carlsberg Res. Commun.* 44:47–63
219. Wagner, G., Wüthrich, K. 1979. *J. Mol. Biol.* 134:75–94
220. Schreier, A. A., Baldwin, R. L. 1976. *J. Mol. Biol.* 105:409–26
221. Schreier, A. A., Baldwin, R. L. 1977. *Biochemistry* 16:4203–9
222. Furie, B., Schechter, A. N., Sachs, D. H., Anfinsen, C. B. 1975. *J. Mol. Biol.* 92:497–506
223. Schmid, F. X., Baldwin, R. L. 1979. *J. Mol. Biol.* 133:285–87
224. Dolgikh, D. A., Gilmanshin, R. I., Brazhnikov, E. V., Bychkova, V. E., Semisotnov, G. V., Venyaminov, S. Y., Ptitsyn, O. B. 1981. *FEBS Lett.* 136:311–15
225. Ptitsyn, O. B. 1972. *Pure Appl. Chem.* 31:227–44