Biochem. J. (1990) 270, 1–16 (Printed in Great Britain)

REVIEW ARTICLE

Protein folding

Thomas E. CREIGHTON
MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, U.K.

1. THE PROTEIN FOLDING PROBLEM

To be biologically active, all proteins must adopt specific folded three-dimensional structures. Yet the genetic information for the protein specifies only the primary structure, the linear sequence of amino acids in the polypeptide backbone. Most purified proteins can spontaneously refold *in vitro* after being completely unfolded, so the three-dimensional structure must be determined by the primary structure. How this occurs has come to be known as 'the protein folding problem'.

The problem can be broken down into three different but related questions: (1) By what kinetic process or pathway does the protein adopt its native and biologically-active folded conformation? (2) What is the physical basis of the stability of folded conformations? (3) Why does the amino acid sequence determine one particular folding process and resultant three-dimensional structure, instead of some other?

The ultimate goal, to be able to predict the native folded three-dimensional structure from just the amino acid sequence, should be possible given a full understanding of the above three questions. Although considerable advances have been made in recent years, notably the successful design *de novo* of simple folded proteins [1–3], attainment of the ultimate goal is not imminent, for there appear to be no simple answers to the above questions.

The purpose of this review is to describe how our understanding of the actual folding process has improved in recent years and to put its various aspects into perspective. It will deal solely with water-soluble, globular proteins and will concentrate on small, single-domain proteins. Earlier work, and that on large, multi-domain proteins, has been reviewed extensively elsewhere [4–14]. Limitations of space prevent discussion of recent exciting studies on protein folding *in vivo* [15–17] and on protein catalysts of folding [18,19].

2. THE STABLE CONFORMATIONAL STATES OF PROTEINS

Different protein conformations differ only in the angle of rotation about the bonds of the backbone and amino acid side-chains, except that they may also differ in covalent disulphide bonds, which are unique (sections 5.A and 5.B). The starting and end points for virtually all studies of protein folding are those conformational states that are stable at equilibrium, and these are the only conformations that can be characterized in detail.

A. The native, fully folded state (N)

The native conformations of proteins are known in great detail from the structures determined by X-ray crystallography and, more recently, by n.m.r. [20–24]. A description of these complex three-dimensional structures is beyond the scope of this article, but is available elsewhere [25–28].

Over 100 substantially different protein three-dimensional folding motifs are known, and proteins with non-homologous amino acid sequences usually have different conformations. In contrast, homologous proteins invariably have essentially the same folded conformation, even if their amino acid similarities are minimal [29,30]. The most conspicuously similar aspects of homologous structures are the general conservation of the non-polar character of the side-chains that comprise the folded interior, plus the general prevalence of hydrophilic side-chains at the surface [31]. How much alteration is necessary before a protein no longer folds to its normal conformation, either remaining unfolded or adopting a new conformation, is not certain. Proteins have been found to be surprisingly adaptive to mutations that would be expected to be disruptive, but the hydrophobic core seems to be the most critical aspect for stability of the normal folded state [32–34].

Folded proteins demonstrate varying degrees of flexibility [35], which is of direct relevance to protein folding, in that it reflects the free energy constraints on unfolding and refolding (section 4). Flexibility is greatest at the protein surface, where some side-chains and a few loops have alternative conformations or no particular conformation that is energetically preferred [36]. Although flexibility is least in the interior, even there side-chain rotations occur, and most tyrosine and phenylalanine side-chain aromatic rings are flipping by 180° on the millisecond time-scale [37]. Nevertheless, the basic architecture of the protein generally stays relatively close to the average structure determined by X-ray crystallography or by n.m.r.; the greatest plasticity of conformation is exhibited by small proteins [38]. Certainly no protein is known to adopt alternative fully folded conformations. For example, a protein invariably has the same three-dimensional conformation, differing only in surface side-chains and loops, when crystallized in different ways and into different crystal lattices. The only substantial conformational changes that occur within known protein domains result from proteolysis, the most spectacular example being α_1-antitrypsin [39]. Otherwise, substantial conformation changes in a protein are almost invariably limited to alterations of the relative dispositions of rigid domains or subunits [40]. The folded conformation of a domain is apparently in a relatively narrow free energy minimum, and substantial perturbations of that folded conformation require a significant increase in free energy.

B. The unfolded state (U)

The ideal unfolded protein is the random coil, in which the rotation angle about each bond of the backbone and side-chains is independent of that of bonds distant in the sequence, and where all conformations have comparable free energies, except when atoms of the polypeptide chain come into too close proximity. Steric repulsions are significant between atoms close in the covalent structure, and place limitations on the local

Abbreviations used: BPTI, bovine pancreatic trypsin inhibitor; CI, compact intermediate; GdmCl, guanidinium chloride; N, native, fully folded state of a protein; U, unfolded state of a protein.

flexibility; they would also occur in a fraction of the totally random conformations between atoms distant in the covalent structure and thereby exclude these conformations (the 'excluded volume effect'). In spite of these restrictions on the conformational flexibility of a real random coil, there are very many conformations possible with even a small protein. It would therefore be impossible for a fully unfolded protein to encounter on a finite time-scale all its possible conformations [41]. Also, each molecule in a typical experimental sample of a fully unfolded protein (likely to contain no more than 10^{18} molecules) will probably have a unique conformation at any instant of time. Consequently, the initial stages of folding must be nearly random, but the native conformation is unlikely to be found by a totally random process.

Unfolded proteins in strong denaturants, such as 6 M-GdmCl or 8 M-urea, and disordered polypeptide copolymers, have been demonstrated by Tanford [42] and by Flory [43,44] to have the average hydrodynamic properties expected of random coil polypeptides. There is a wide variety of evidence, however, suggesting that unfolded proteins are not true random coils under other conditions, such as extremes of pH or temperature in the absence of denaturants [45–48]. This is perhaps not too surprising, for in a truly random coil the energetics of interactions between different parts of the polypeptide chain must be exactly balanced by interactions with the solvent. This is virtually impossible with a polypeptide chain composed of 20 different amino acid side-chains, with a diversity of chemical properties. Where interactions between different parts of the polypeptide chain are energetically favoured over those with solvent, the polypeptide chain will tend to be more compact and less disordered than expected for a random coil. The opposite will occur where there are especially favourable interactions between solvent and polypeptide. Nevertheless, unfolded states produced under different unfolding conditions, which often have different physical properties, are indistinguishable thermodynamically, so they are probably different sub-sets of the truly random spectrum of non-native conformations [47–49]. Most importantly, unfolded proteins do not generally contain co-operative folded structures (see section 3.B).

C. The compact intermediate (CI), so-called 'molten globule', state

A variety of proteins have been observed under certain conditions to exist in stable conformations that are neither fully folded nor fully unfolded. These conformations have sufficient similarities to suggest that they are different manifestations of a third stable conformational state [50,51]. The most common properties are: (1) The overall dimensions of the polypeptide chain are much less than those of a random coil and only marginally greater than those of the fully folded state. (2) The average content of secondary structure is similar to that of the folded state. (3) The interior side-chains are in homogeneous surroundings, in contrast to the asymmetric environments they have in the fully folded state. (4) Many interior amide groups exchange hydrogen atoms with the solvent more rapidly than in the folded state, but more slowly than in the fully unfolded state. (5) Its enthalpy is very nearly the same as that of the fully unfolded state, substantially different from that of the native state. (6) Interconversions with the fully unfolded state are rapid and nonco-operative, but slow and co-operative with the fully folded state (Section 3.B).

If these observations are applicable to a homogeneous structure, they suggest a collapsed molecule with native-like secondary structure and a liquid-like interior, i.e. a 'molten globule'. Detailed studies of one such protein, α-lactalbumin, however, have demonstrated that portions of the hydrophobic interior are in relatively stable, well-ordered three-dimensional conformations, with amide groups highly protected from hydrogen exchange, whereas other parts are much less structured [52]. Viewing this state as a homogeneous molten globule may therefore be misleading, so the term 'compact intermediate' (CI) will be used here [12]. The CI state appears to be the preferred conformational state of the unfolded protein under refolding conditions, where it is usually only transient (see section 4.C.3). There may be a continuum of unfolded conformations, with the CI state at one extreme, fully folded at the other.

3. GENERAL PROPERTIES OF PROTEIN FOLDING TRANSITIONS

The native conformational states of proteins may usually be unfolded reversibly by adding denaturants, increasing or decreasing the temperature, varying the pH, applying high pressures, or cleaving disulphide bonds. At equilibrium, unfolding transitions of single-domain proteins are usually two-state:

$$N \leftrightarrow U \quad (1)$$

with only the fully folded (N) and unfolded (U) states populated [49], although the CI state may be a subset of U (Fig. 1). Partially-folded conformations, with thermodynamic properties distinctly different from either U or N, are energetically unstable relative to either U or N under all conditions. The most exceptional and intriguing behaviour is demonstrated by certain proteins with eight-stranded β-barrel $(\alpha/\beta)_8$ conformations, where stable folded conformations of unknown nature are adopted by segments of the polypeptide chain that comprise only part of the $(\alpha/\beta)_8$ domain [53].

Multi-domain proteins usually unfold step-wise, with the domains unfolding individually [54], either independently or with varying degrees of interactions between them [55,56]. Multi-subunit proteins usually dissociate first, then the subunits unfold, unless domains are on the periphery of the aggregate where they can unfold independently [13,14]. The primary consideration, therefore, is the unfolding/refolding transition of a single protein domain, which will be the subject of the remainder of this review.

A. Stability of the folded state

Understanding the physical basis of stability of the folded state is crucial for understanding how such a conformation can be acquired. The general absence at equilibrium of partially-folded states in two-state unfolding transitions makes it relatively easy to measure the equilibrium constant and free energy of folding within the transition region, where both N and U are populated:

$$K_{eq.} = [N]/[U] \quad (2)$$

$$\Delta G^0{}_N = G^0{}_N - G^0{}_U = RT \ln K_{eq.} \quad (3)$$

The net stability of the folded state outside the transition region must be determined by extrapolation. Within the transition region of two-state unfolding, the value of $\Delta G^0{}_N$ is usually observed to be linearly dependent upon the concentration of urea or GdmCl. This linear dependence is usually extrapolated to estimate the value of $\Delta G^0{}_N$ in the absence of denaturant [57], even though its physical basis is not understood [58,59].

The folded states of proteins are only marginally more stable than the fully unfolded state, even under optimal conditions [49]. Typical values of $\Delta G^0{}_N$ for small natural proteins are -5 to -10 kcal/mol (-20 to -40 kJ/mol), so the equilibrium constant between the N and U states would have a value in the region of 10^4–10^7. One consequence of this is that most folded proteins must be spontaneously unfolding completely under all

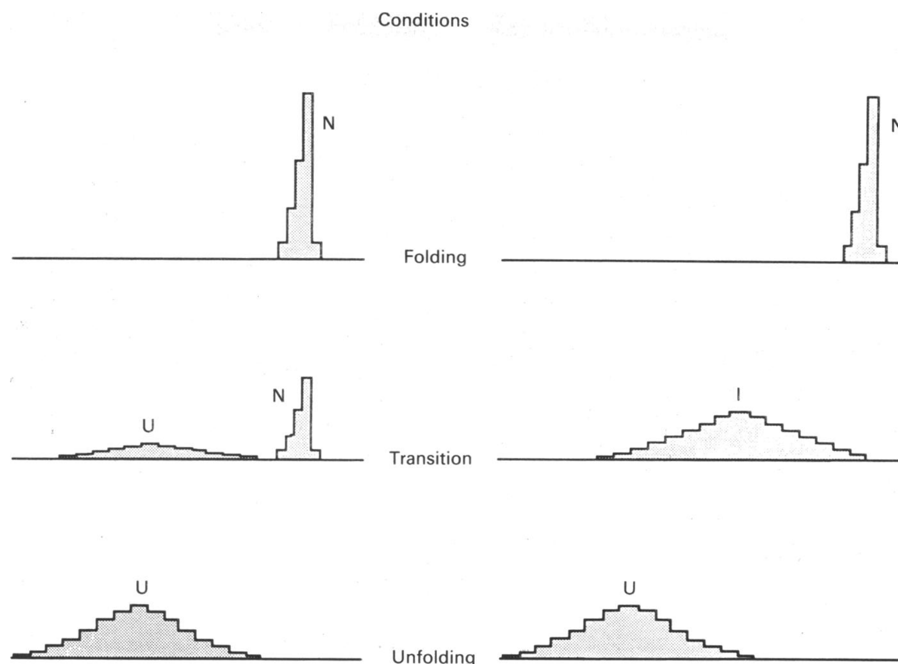


Fig. 1. Illustration of a co-operative, two-state folding transition (left), as normally observed with single-domain proteins, compared with a nonco-operative, multi-state transition (right)

Histograms of the numbers of molecules (vertical axis) with different degrees of folding (horizontal axis) are illustrated for conditions favouring folding (top) and unfolding (bottom), and for the transition region (middle). The folded state, N, is represented as a narrow distribution of folded conformations, the unfolded state, U, as a broad distribution of less compact conformations. Within the transition region, a two-state transition will have two distinct populations in equilibrium, similar to those at the two limiting conditions of folding and unfolding. In contrast, a nonco-operative transition will have the distribution of the entire population shifted so that most molecules have intermediate, partially-folded conformations (I) within the transition region.

conditions. If the protein refolds spontaneously with a rate constant of 1 s^{-1}, its rate of spontaneous total unfolding under the same conditions will be 10^{-4}–10^{-7} s^{-1}, i.e. half-times of 2 h to 80 days. This spontaneous unfolding will normally be only transient, because the protein will promptly refold.

Natural proteins apparently have not been selected for maximum stability, for a synthetic protein designed empirically is much more stable, $\Delta G^0_N = -22.5$ kcal/mol (-94 kJ/mol) [1]. Natural proteins appear to require some flexibility for their function, or possibly to be able to fold into the native conformation quickly, both of which would be hampered by too stable a final folded conformation (see section 5.B)

The thermodynamics of folding transitions have been thoroughly characterized calorimetrically by Privalov [49,54,60]. The enthalpies (ΔH) and entropies (ΔS) of unfolding are very temperature-dependent (Fig. 2), because the heat capacity of the unfolded state is significantly greater than that of the folded state. The usual interpretation is that the heat capacity difference results primarily from the temperature-dependent ordering of water molecules around the non-polar portions of the protein molecules, more of which are solvent accessible in the unfolded state [49,60–62], although other factors may contribute.

The large heat capacity change upon protein unfolding causes there to be a temperature at which stability of the folded state is at a maximum [49,60,63]. Measured by free energy, the maximum occurs where $\Delta S = 0$, while that measured by the equilibrium constant occurs where $\Delta H = 0$ [64]. These maximum stabilities can occur at quite different temperatures, but both are used. In any case, the stability of the folded state decreases at both higher and lower temperatures. Proteins may almost always be unfolded by raising the temperature sufficiently, but unfolding at low temperatures can be observed only under circumstances where it occurs within an accessible temperature range. Low-temperature unfolding has the opposite thermodynamic characteristics of high-temperature, except for the same heat capacity change, but is not fundamentally different [47,65,66]. The unfolded state is the same thermodynamically, and the opposite thermodynamic parameters are simply a consequence of the large heat capacity change.

There seems little doubt that the hydrophobic interaction is a major contributor to the stability of the folded state, essentially as proposed by Kauzmann [67], although the phenomenon is complex and there is considerable controversy [60,64,68–70]. Nevertheless, there must also be other significant factors stabilizing proteins, for the enthalpy of folding is more negative than that expected from just the hydrophobic effect [27,64,71]. The other types of interactions present within folded proteins, such as hydrogen bonds, van der Waals' interactions, and electrostatic interactions, have traditionally been assumed to provide no net contribution to overall stability of the folded state, because comparable interactions should be made between the unfolded state and the solvent. This conclusion, however, neglects the intramolecular nature of the interactions within the folded state, as opposed to the intermolecular interactions between solvent and protein [27,71]. Interactions within the folded state can have substantially lower free energies than those between solvent and protein for simple entropic reasons [72]. This entropic effect can also be reflected in the enthalpy if the intramolecular interactions are as a consequence also more favourable enthalpically. For example, most hydrogen bonds within water [70] and between protein and water are usually present only a fraction of the time [27], whereas those within

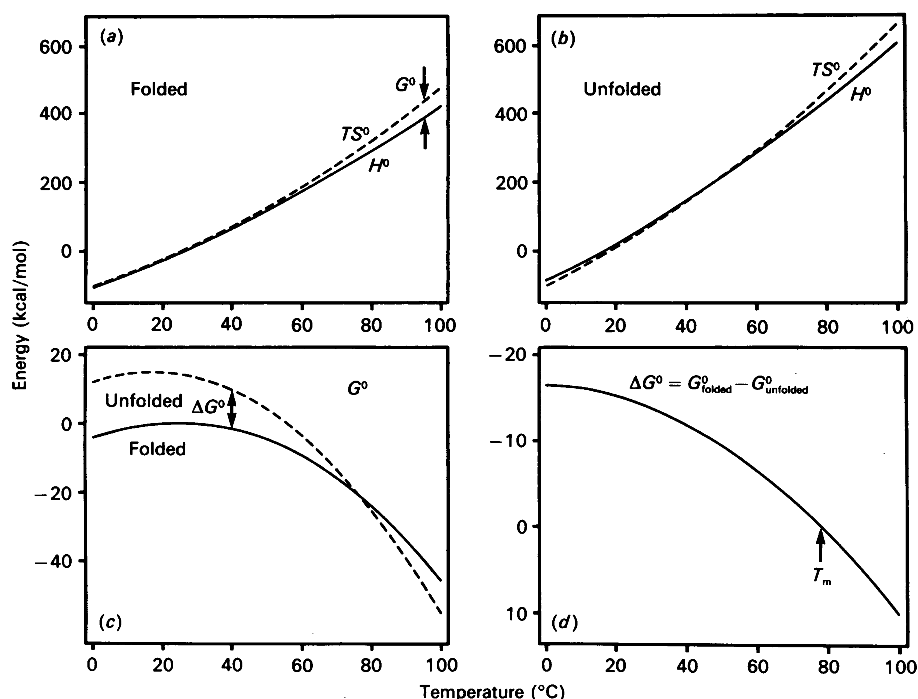


Fig. 2. Thermodynamic parameters for the folded and unfolded forms of lysozyme in water at pH 7.0 and various temperatures

The enthalpic (H^0) and entropic (TS^0) contributions to the free energies are plotted as a function of temperature in (*a*) for the folded state and (*b*) for the unfolded state. Both vary with temperature because of the substantial heat capacities of both states. The Gibbs free energies (G^0) of both states are the differences in the enthalpic and entropic contributions, as illustrated in (*a*). The values of G^0 for both states are plotted in (*c*) as a function of temperature. The net stability of the folded state (ΔG^0) is the difference between the free energies of the two states, as indicated in (*c*), and is plotted in (*d*). The melting temperature of native lysozyme is designated by T_m, that temperature where $\Delta G^0 = 0$. Note the change in free energy scales from (*a*) and (*b*) to (*c*) and to (*d*); the final ΔG^0 is a very small difference between the individual enthalpy and entropy contributions. The curvature of the plot of ΔG^0 (*d*) results from the substantial heat capacity difference between the folded and unfolded states. Constructed with data taken from [178].

folded proteins are present essentially all of the time [73]; the latter should consequently have the more negative enthalpy. Van der Waals' interactions within the close-packed protein interior should be substantially greater than those between the protein and the solvent; they should have correspondingly lower enthalpies, analogous to the enthalpy of fusion when liquids crystallize [74].

It therefore seems reasonable to conclude that most of the interactions within the folded protein are more favourable energetically, in both enthalpy and free energy, than the corresponding interactions of the unfolded state. They should therefore contribute to the net stability of the folded state, and it is not surprising that hydrogen bonds and salt-bridges have been found to do so [75–77]. Nevertheless, the hydrophobic interaction is probably the major stabilizing factor.

Even though the folded state is stabilized by many interactions, it has only marginal net stability, because there are compensating factors that stabilize the unfolded state. The most substantial of these is its large favourable conformational entropy. Consequently, the net stability of the folded state, $\Delta G^0{}_N$, is a relatively small difference between the substantial compensating interactions stabilizing the folded and unfolded states (Fig. 2).

B. Co-operativity of folding

The two-state nature of protein folding transitions indicates that folding is a co-operative process [49]. Little happens to the fully folded state prior to complete unfolding; e.g. it is not detectably perturbed by varying the temperature [78,79]. Effects of denaturants are observed [80] but may arise because of direct interactions with the protein [81]. Once unfolding of a domain is initiated, it proceeds to completion. The interactions stabilizing the folded conformation must be co-operative; breaking one or more of the interactions must destabilize the others so that the free energy increases and the folded conformation becomes unstable.

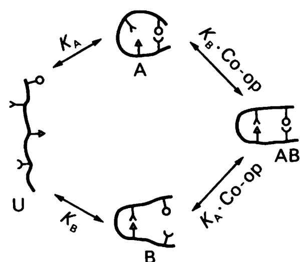
Co-operativity of folding transitions is usually inferred from one or more characteristics. (1) Intermediate conformations should not be populated to a substantial extent, so all probes of unfolding should follow the same unfolding curve. (2) The enthalpy change measured calorimetrically should be the same as that obtained by van't Hoff analysis [49], where the dependence of the equilibrium constant on temperature is measured (see Fig. 5). (3) The stabilities of the interactions within the fully folded state should be greater than in all other conformations, making the fully folded state a unique conformation. (4) The unfolding transition should be much more abrupt than expected from the disruption of a single interaction. The best example is pH titration; the acid-induced unfolding of proteins usually occurs over a smaller pH range than does titration of a normal ionizing group. This arises from the presence of multiple ionizable groups within the protein interior, especially histidine residues, which can ionize only after the protein unfolds [82]. Unfolding consequently occurs abruptly, and all the groups ionize in concert.

Co-operativity of denaturant-induced unfolding is often inferred if it occurs over a limited range of denaturant concentration, to give a sigmoidal unfolding curve. This, however, only indicates that the equilibrium constant for folding in the absence of denaturant is sufficiently large that the proportion of

unfolded molecules is negligible initially. There is no equivalent of the Henderson–Hasselbach equation to predict the properties expected for disruption of a single interaction, so such a sigmoidal unfolding curve is not necessarily evidence for co-operativity.

There are probably two major reasons for the co-operativity of folding transitions. The first concerns unfavourable interactions in the partially-folded states that are not present in either the fully folded or unfolded states. Plausible examples would be the increase in free energy produced by breaking an internal hydrogen bond without supplying comparable hydrogen-bonding partners to the acceptor or donor and by pulling apart two non-polar surfaces sufficiently far that the van der Waals' interactions are greatly diminished, but without gaining comparable interactions with other surfaces or with the solvent. Any such conformational strain present in partially-folded structures, but not in U or N, should contribute to the co-operativity of folding, but not to the net stability of the folded state.

The second reason for co-operativity is likely to be the entropic co-operativity of the folded state, which is necessary to account for its stabilization by intrinsically weak interactions [27,71]. This co-operativity arises from the simultaneous presence of many interactions within a single conformation; their total contribution to stability may be much greater than that of the sum of the individual interactions. This can be illustrated by considering a simple conformation with only two stabilizing interactions, A and B, which may be hydrogen bonds, van der Waals' interactions, etc.:



U includes all the (unfolded) conformations with neither interaction, and the other species are depicted by the interactions that are present. K_A and K_B are the equilibrium constants for formation of interactions A and B, respectively, in the unfolded state. Values for K_A and K_B for typical protein groups distant in the polypeptide chain are expected to be within the range 10^{-6}–10^{-1} [27,71]. In other words, individual hydrogen bonds, salt bridges, van der Waals' interactions, etc. are not expected to be stable and present at high frequency within an unfolded polypeptide chain, unless in particularly favourable proximity in the covalent structure.

Entropic co-operativity arises between two or more interactions if the presence of one brings the other potentially interacting groups into more favourable proximity and orientation, when '*Co-op*' > 1. If the interactions are independent, *Co-op* = 1, whereas *Co-op* < 1 if one interaction interferes with the other and there is anti-co-operativity. A single value of *Co-op* pertains to both interactions, so the magnitude of the effect of interaction A on B must be the same as that of B on A. Each interaction affects the other to the same extent. The importance of entropic co-operativity in the folded state when '*Co-op*' > 1 is that it contributes to the free energy of stabilization of the folded state. In the simple example above, the equilibrium constant between the 'folded' state AB and U will be given by

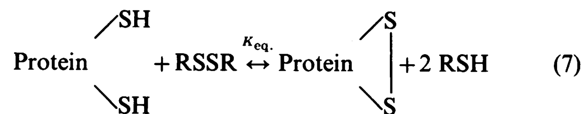
$$K_{eq.} = \frac{[AB]}{[U]} = K_A K_B \, Co\text{-}op \qquad (5)$$

In terms of the free energies of the interactions (e.g., $\Delta G^0_a = -RT \ln K_A$), the free energy of the folded state, G^0_{AB}, relative to that of the unfolded state, G^0_U, is given by:

$$G^0_{AB} - G^0_U = \Delta G^0_a + \Delta G^0_b - RT \ln Co\text{-}op \qquad (6)$$

Co-operativity between three or more interactions can be examined by extending the above conformational equilibria. Different *Co-op* values will be relevant to different pairs of interactions, but the principles are the same. The *Co-op* factors for each successive interaction are all present in the expression for the final equilibrium constant between the initial, unfolded state and the fully folded state with all the interactions present simultaneously. Consequently, each interaction within the final folded conformation should be more stable than in the unfolded protein. With sufficient number of co-operative interactions, even if intrinsically weak, the free energy of the folded state will become lower than that of the unfolded protein. The fully folded state is then stable, and the folding transition is co-operative [71].

This explanation of the co-operativity of protein folding is based upon experimental measurements of the stabilities of the disulphide bonds of BPTI during folding [9,71]. Most studies of the contributions to stability of individual interactions use site-directed mutagenesis to remove one interacting group and measure the change in stability [83–85], but this type of analysis is complicated by effects of the mutations on the folded conformation [84] and, perhaps, on the unfolded state [86]. Removing a group by mutagenesis is unlikely to leave a void in a folded conformation, so conformational rearrangements are almost certain to occur. In contrast, the disulphide interaction between cysteine residues is unique in that its strength can be varied experimentally through the ratio of thiol (RSH) and disulphide (RSSR) reagents in the solution:



$$\text{Protein}\begin{matrix}\diagup SH\\ \diagdown SH\end{matrix} + RSSR \overset{K_{eq.}}{\leftrightarrow} \text{Protein}\begin{matrix}\diagup S\\ |\\ \diagdown S\end{matrix} + 2\,RSH \qquad (7)$$

The fraction of molecules with the disulphide bond can be determined after trapping them, so a wide range of relative stabilities of individual protein disulphides may be measured experimentally [87]. During folding and formation of the three disulphides of BPTI, the stabilities of the disulphides increase by factors of 10^3–10^7, due to co-operativity between disulphide bond formation and the protein conformation. The increased stabilities in the folded conformation are due not only to the undoubtedly substantial entropic factors, but also to contributions from unfavourable energetics unique to the intermediate states lacking a disulphide bond, from cysteine thiol groups being placed within the interior of the folded protein, where a disulphide normally exists. The increased disulphide stability in the folded conformation can be attributed primarily to entropic factors in the BPTI intermediate without the weakest disulphide, between cysteines 14 and 38. These two cysteine residues are on the surface, so unfavourable effects of having two thiol groups there are probably minimal. Entropic co-operativity should also be least on the surface, so the weak surface disulphide between cysteines 14 and 38 gives a minimum estimate of 10^3 for the entropic co-operativity contribution of each disulphide to the stability of folded BPTI.

The co-operativity of folded conformations will vary, as will the contributions of individual interactions [88–90]. Consequently, there is probably no standard value for the net stability within folded proteins of a hydrogen bond, or any other type of interaction, only a normal range in which it may occur. Nevertheless, interactions between groups in the most rigid parts of the protein interior are expected and observed to make the

greatest contributions to stabilization of the folded state [91,92]. A more rigorous and satisfactory description of protein stability requires the ability to calculate the entropies and free energies of protein conformations.

4. KINETICS OF PROTEIN UNFOLDING AND REFOLDING

Proteins are believed to fold relatively quickly because they follow a defined mechanism and pathway of folding and do not rely on random fluctuations to encounter by chance the correct folded conformation [41]. That the potential for existing in a stable folded conformation is not sufficient to guarantee a pathway to reach it is indicated by proteins that cannot refold [93–96]. These examples are proteins that normally fold as a precursor or after translocation through a membrane; presumably the normal folding pathway is not accessible after unfolding of the mature protein because part of the original protein, or some other aspect of the biosynthetic machinery, is missing. Further indirect evidence comes from mutations that block folding to a dramatic extent, but do not alter the stability of the folded state [85,97]. Not surprisingly, the experimentally determined pathways of folding are far from random (Sections 4.C and 5).

Not all aspects of the primary structure are always required for folding. At least some proteins with circularly permuted sequences can fold at nearly normal rates to the same folded conformation [98,99]. Clearly, the termini of neither the original nor the permuted primary structures are crucial, so folding need not proceed from the *N*-terminus of the polypeptide chain. Similarly, the linear arrangement of all the amino acids in the polypeptide chain is not crucial. Whether this result would be obtained with permutations in which termini are inserted between elements important for the initial stages of folding remains to be tested experimentally.

A. Kinetic analysis of complex reactions

To determine the mechanism and pathway of unfolding and refolding, the intermediates that define and direct the pathway must be identified, but these are usually unstable thermodynamically. They might be detectable as kinetic intermediates, but only if they occur on the pathway before the rate-limiting step* and if their free energies are comparable to, or lower than, that of the initial state. Any intermediates with more positive free energies will not be populated even transiently.

It is probably unreasonable to expect the folding intermediates that are crucial for increasing the rate of folding to be highly populated under physiological conditions, even kinetically. Consider an intermediate state, I, that is necessary for a defined sequential pathway, of the type:

$$U \leftrightarrow I \overset{\text{slow}}{\leftrightarrow} N \qquad (8)$$

Increasing the stability of this intermediate relative to U, with the rate-determining step constant, will increase the rate of the folding reaction only until the free energy of I is comparable to that of U (Fig. 3); making the intermediate more stable than U will produce no increase in rate. Therefore, an obligatory intermediate is unlikely to be populated to an extent greater than the unfolded protein under physiological conditions.

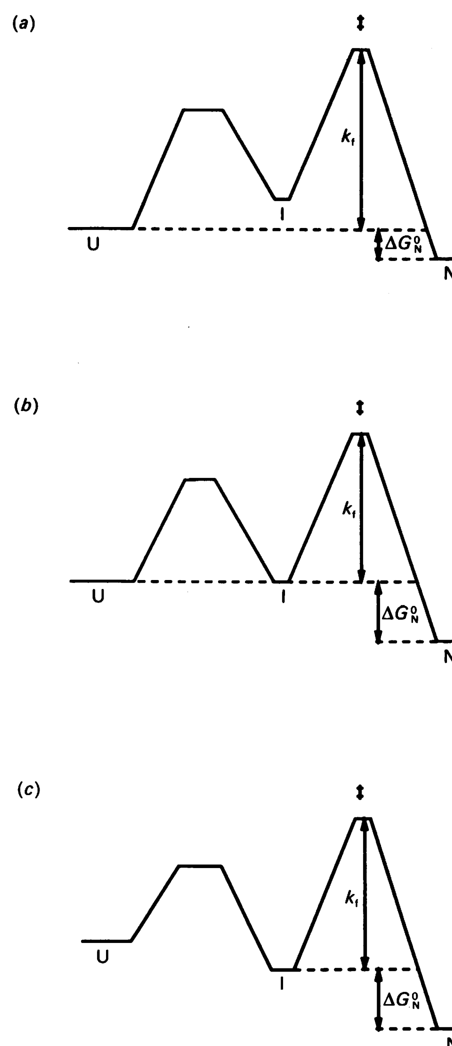
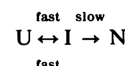


Fig. 3. Illustration that obligatory intermediates to promote the rate of protein folding would not be expected to be stable under physiological folding conditions

Simple free energy diagrams are illustrated for the reaction



$$U \underset{\text{fast}}{\overset{\text{fast}}{\leftrightarrow}} I \overset{\text{slow}}{\rightarrow} N$$

where I is an obligatory intermediate. The relative free energies of I, N, and the transition state (‡) remain unchanged, while that of U varies. The relevant free energy differences that determine the rate of folding (k_f) and the net stability of the folded conformation ($\Delta G^0{}_N$) are indicated. With an unstable I (*a*), increasing its stability relative to U causes the stability of N to increase and the rate of folding to increase (i.e., the free energy barrier decreases), until I has the same stability as U (*b*). With I more stable than U (*c*), the rate of folding and the stability of N no longer increase; I merely becomes the predominant form of the unfolded protein under folding conditions.

With a simple one-step reaction, a single kinetic phase, characterized by a single rate constant, k, is expected:

$$\text{fraction folded conformation} = 1 - \exp(-kt) \qquad (9)$$

More complex behaviour would be observed either if there were

* The rate limiting step is used here as defined by Gold [179]: 'the earliest step for which the forward chemical flux ... is practically equal to the rate of formation of the final product'. The over-all rate of the reaction will be given by the microscopic rate constant for this step multiplied by the equilibrium constants for any unfavourable pre-equilibria steps that precede it. The rate-limiting step will be that with the highest free energy barrier along the reaction co-ordinate, unless there is a stable intermediate. In this case, the rate-limiting step is that with the greatest free energy difference between its transition state and the lowest free energy intermediate preceding it [180].

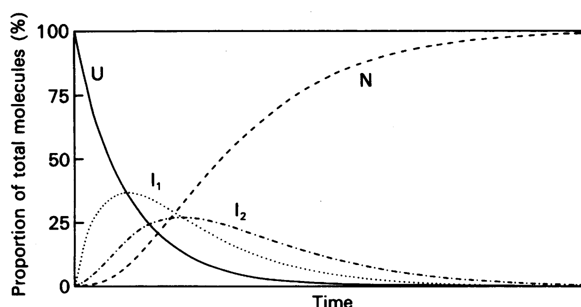


Fig. 4. Kinetics of an obligatorily sequential pathway with multiple rate-limiting steps

The progress of the reaction $U \rightarrow I_1 \rightarrow I_2 \rightarrow N$ was simulated with each step having the same rate. The proportion of molecules in each species is plotted as a function of time. Note that U disappears and I_1 appears without a lag, but that I_2 and N appear only after lag periods. That for I_2 corresponds to the time during which precursor I_1 is accumulating, while that for N arises from the need to build up the concentration of I_2 also. The lag period for N is greater than that for I_2 because of the greater number of preceding slow steps.

multiple rate-limiting steps or if the starting material were heterogeneous, with different populations having different rates of reaction. Discriminating between these possibilities is not always straightforward, but it is too often assumed that one kinetic phase of a folding reaction represents formation of an obligatory intermediate, I, a second its conversion to N:

$$U \underset{k_{-1}}{\overset{k_1}{\leftrightarrow}} I \overset{k_2}{\rightarrow} N \qquad (10)$$

If this were the case, there must be a lag period in the appearance of N, of approximate magnitude $(k_2+k_1)^{-1}$, during which the steady-state concentration of I is generated (Fig. 4). This effect is cumulative, so the magnitude of the lag period in formation of the final folded conformation should be correspondingly longer with increasing number of obligatory, sequential intermediates along a pathway. Very few claims of obligatory intermediates in protein folding are supported by the observation of an appropriate lag period.

Protein folding is special because of the great conformational heterogeneity of the unfolded state, in which every molecule of a typical population is likely to have a unique conformation at each instant of time. How is this heterogeneity apparent in the kinetics of refolding? Does each molecule refold with a unique rate, determined by its conformation at time zero [100], or do molecules somehow fold by a common mechanism and rate? If each molecule does not fold uniquely, how do different molecules manage to follow the same rate-limiting step?

It is clearly unrealistic to expect to elucidate all the details of a complex reaction like protein folding. Although it occurs much more rapidly, on the second to minute time-scale, than expected for a random search, this time is sufficiently long for each molecule to undergo some 10^{11} to 10^{13} conformational changes. Because each molecule is starting out with a different conformation, it might be feasible to determine only at what stage different molecules start to follow the same pathway. At best it may be possible only to characterize the slowest transitions and the conformations and energetics of the most stable intermediates, to identify the overall rate-limiting step, and to characterize the transition state.

B. Kinetics of protein unfolding

The kinetics of unfolding are important for characterizing the overall folding transition. Unfolding is almost universally

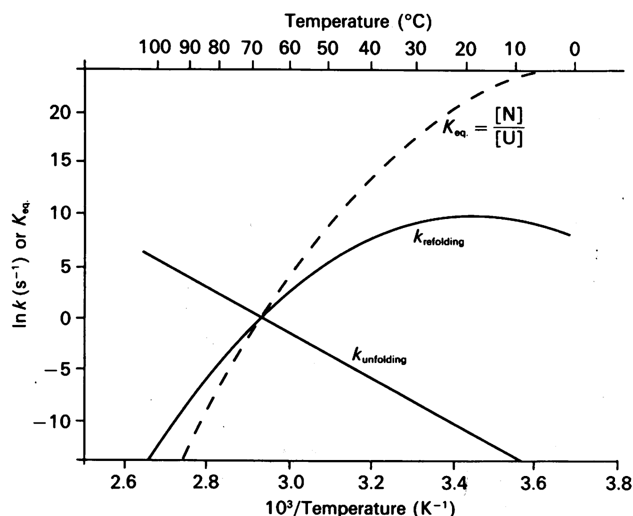


Fig. 5. Typical temperature-dependence of the rates and equilibria of protein folding transitions not involving intrinsically slow isomerizations

The natural logarithms of the rate constants for unfolding and refolding are plotted as a function of 1/temperature, in an Arrhenius plot. A similar plot of the equilibrium constant ($K_{eq.}$) between the folded (N) and unfolded (U) states is a Van't Hoff plot. The curvature of the Van't Hoff plot is due to the greater apparent heat capacity of U than of N. The linear Arrhenius plot for the rate of unfolding indicates that the transition state has the same heat capacity as N. The greater heat capacity of U is reflected entirely in the curvature of the Arrhenius plot for the rate of refolding, because $\ln K_{eq.} = \ln k_{refolding} - \ln k_{unfolding}$. The data used to construct this diagram are for hen egg-white lysozyme at pH 3. The data for stability were from [178], for rates of folding from [146], extrapolated to the absence of GdmCl. Although $k_{unfolding} = k_{refolding}$ at $K_{eq.} = 1$, it is a coincidence that the rate constants had the value 1 s^{-1} at this temperature, so that all three curves intersect at a common point.

observed to be an all-or-none process, with little or no partial unfolding preceding it. Upon placing at time zero a native, covalently homogeneous protein into unfolding conditions, unfolding almost always occurs with a single kinetic phase and a single rate constant. There is no lag period, and all probes of unfolding generally give the same rate constant. Therefore, there is a single rate-limiting step in unfolding, and all the folded molecules have the same probability of unfolding. This is not surprising in view of the generally homogeneous nature of the folded state.

The rate of unfolding usually changes uniformly with variation of the unfolding conditions. In particular, Arrhenius plots of unfolding rates (Fig. 5) are generally linear, suggesting that the mechanism of unfolding is not changing. There appears to be a single transition state for unfolding.

C. Kinetics of protein refolding

Kinetic complexities are encountered almost universally in protein refolding [6], which usually result from conformational heterogeneity of the unfolded state, with slow- and fast-refolding molecules:

$$U_S \overset{slow}{\leftrightarrow} U_F \overset{fast}{\leftrightarrow} N \qquad (11)$$

In virtually all characterized cases, the heterogeneity arises from *cis–trans* isomerization of peptide bonds preceding proline residues [101].

1. Peptide bond isomerization. The peptide bond is usually planar, due to its partial double-bond character, and can exist in either the *cis* or *trans* isomer. The *trans* form is intrinsically favoured energetically about 10^3-fold over the *cis* form, due to

the latter having unfavourable steric clashes between the C^α atoms and side-chains of neighbouring residues. When the next residue is proline, however, its cyclic side-chain minimizes many of these unfavourable interactions, and the *cis* and *trans* forms have comparable free energies; typically, the *trans* form is favoured only 4-fold. Consequently, *cis* peptide bonds are often found in folded proteins when the next residue is proline, particularly at reverse turns on the surface of the protein [27].

Within a folded protein, a peptide bond is usually *cis* or *trans* in essentially all the molecules, for the folded conformation generally favours one over the other. In a few cases, however, both forms may be evident in the folded state [102,103]. Protein conformational stability and *cis–trans* isomerization of peptide bonds are linked functions, so whatever effect a protein conformation has on stability of the *cis* and *trans* isomers of a peptide bond, the isomer must have exactly the same effect on the stability of the folded conformation.

When a protein is unfolded, the constraints favouring one form over the other are released, and an equilibrium between *cis* and *trans* isomers at each peptide bond is attained. When the protein is refolded, a fraction of the molecules, U_F, will have all the necessary peptide bonds as the correct isomer while the others, U_S, have one or more as an incorrect isomer.

Cis-trans isomerization of proline peptide bonds is intrinsically slow, with a half-time at 0 °C of 20 min, decreasing by a factor of about 3.3 for each 10 K rise in temperature [101]. This is often slow relative to the rate of folding of the U_F molecules, so they refold rapidly, and the other, U_S, molecules fold more slowly. If all the peptide bonds must be of the correct isomer for refolding to occur, the fraction of U_S molecules increases with the number of proline residues, and these molecules refold more slowly [104]. The actual situation is more complex, however, for some proteins can refold to a native-like conformation with an incorrect isomer of one or more peptide bonds. The rates of isomerization can be either increased or decreased by the conformation of the protein [105].

For example, bovine ribonuclease A has four proline residues, two of which have *cis* peptide bonds in the folded conformation. Unfolded ribonuclease A refolds with three different kinetic phases, corresponding to at least three different unfolded species [6,106]. One accounts for 15% of the molecules and refolds within milliseconds under optimal conditions; it is believed to have all correct peptide bond isomers. A second represents 65% of the molecules and refolds on the second time-scale; under conditions strongly favouring folding, it can fold more readily into a native-like conformation with the incorrect peptide bond isomer. The remaining 20% of the molecules refold even more slowly. Although it is agreed that these different unfolded states differ in *cis–trans* isomers of peptide bonds, the identities of the peptide bonds are contentious [107].

Replacing proline residues, especially those with *cis* peptide bonds in the folded state, with other amino acids can abolish the slow refolding transitions [108–110]. This is often complicated, however, by destabilizing effects on the folded conformation, as a *cis* peptide bond is unfavourable without the proline residue.

In view of the rate-limiting significance of *cis–trans* proline peptide bond isomerization for protein folding, it is perhaps not surprising that proteins capable of catalysing this isomerization exist [19,111]. Remarkably, two such proteins are also the physiological receptors for different immunosuppressants, cyclosporin and FK506 [112–115]; the relationship between peptide bond isomerization and immunosuppression is intriguing but unclear.

2. Refolding in the absence of slow peptide bond isomerization. Within a population of unfolded molecules with the same *cis–trans* isomers, the refolded protein generally appears without a significant lag period and with a single rate constant. The absence of an observable lag period indicates that there is a single rate-limiting step in refolding and that all preceding and subsequent steps must be rapid, and probably reversible. Refolding can consequently be simplified to three stages (Fig. 6*c*): (1) the nature of the unfolded protein under refolding conditions, the 'pre-folded' conformation (section 4.C.3); (2) the nature of the

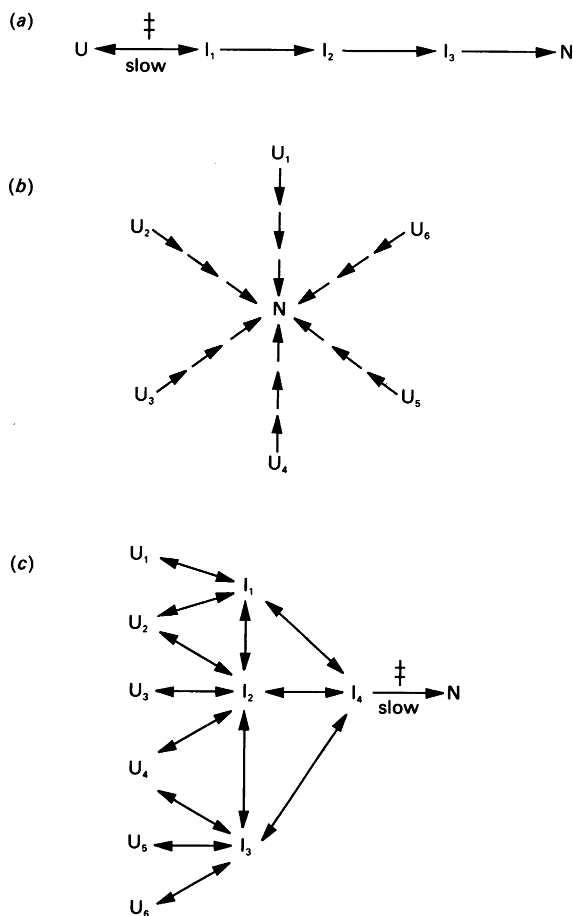


Fig. 6. Examples of models for protein folding, in the absence of intrinsically slow isomerizations

U_i are various unfolded molecules with different conformations at the start of folding, I_i are partially-folded molecules, and N is the fully-folded protein. All kinetic steps indicated by arrows are rapid, except for those labelled 'slow'; ‡ indicates the occurrence of the overall transition state. Single-headed arrows indicate steps that effectively occur only in the indicated direction under conditions strongly favouring folding. (*a*) The nucleation, rapid-growth model [145], in which a nucleation event in the unfolded protein, here indicated as formation of I_1, is the rate-limiting step. The nucleation event is very local and occurs randomly, so it could occur in all the unfolded molecules. Subsequent steps through various intermediates are rapid and essentially irreversible under strongly folding conditions. (*b*) The jig-saw puzzle model [100], in which each unfolded molecule folds by a unique sequence of events. The different pathways only converge at the fully-folded conformation. Each pathway will occur with a unique rate and must be essentially irreversible. (*c*) The model indicated by the experimental data. All the unfolded molecules rapidly equilibrate under folding conditions with a few partially-folded, marginally stable intermediates, which are also in rapid equilibrium. All the molecules pass through a common slow step, which involves going through a transition state that is a distorted form of the native-like conformation. Any intermediates that occur after the rate-limiting step are probably very unstable relative to N.

rate-limiting step and the transition state for folding (section 4.D), and (3) the nature of the folded conformation under refolding conditions, especially its flexibility (section 2.A).

Considering the conformational heterogeneity of the unfolded state (but excluding intrinsically slow isomerizations), it is noteworthy that all the molecules usually fold with the same rate constant. In exceptional cases where more than one kinetic phase is evident in refolding [116], it is also observed in unfolding, which suggests that covalent heterogeneity of the protein is responsible. A single rate constant is consistent with all the molecules folding via the same rate-determining step (Fig. 6*c*), but it only indicates that all the molecules have the same probability of refolding, and they actually fold over a range of times (eqn. 9). It is not inconceivable that a distribution of folding times for a population of molecules folding by different paths might happen to mimic that expected with a homogeneous population and a single rate constant, but it would be unlikely to occur always, under all folding conditions and with different proteins. The alternative explanation that all the molecules are following the same rate-limiting step seems much more plausible.

Folding of all the conformationally heterogeneous unfolded molecules by the same rate-limiting step requires that there be a rapid conformational equilibration prior to the rate-limiting step [117] (Fig. 6*c*). That this occurs is also indicated by the general observation that the rate of refolding depends upon only the final folding conditions, not upon the initial unfolding conditions [118–121]. Proteins unfolded in different ways generally have different average physical properties (section 2.B). Nevertheless, they refold at indistinguishable rates under the same final folding conditions. The rate of folding is determined not by the nature of the initial unfolded protein but by the properties it rapidly adopts when placed under the final folding conditions.

How do all the unfolded molecules equilibrate rapidly prior to refolding, if sampling of all conformations by a random coil would require such a long period of time? The answer undoubtedly is that an unfolded protein under refolding conditions is not a random coil, but adopts a limited set of energetically-favoured non-random conformations, the pre-folded state, as described in the following section. In this way, all the molecules converge to follow a common subsequent pathway and the same rate-limiting step, in contrast to the proposal that each protein molecule folds by a unique pathway [100] (Fig. 6*b*). Most experimental indications of multiple folding pathways arise from different populations separated by the high energy barrier of *cis–trans* peptide bond isomerization [122].

The rate of refolding generally varies with temperature in a complex manner, giving a non-linear Arrhenius plot (Fig. 5). At low temperatures, the rate of refolding increases with increasing temperature, as do most chemical reactions. The increase in rate diminishes, however, and the rate reaches a maximum and then decreases dramatically at high temperatures. This temperature-dependence is unusual for chemical reactions, but might be expected for a complex reaction like protein folding that should depend upon the presence of meta-stable, partially-folded intermediates. Such meta-stable intermediates would be destabilized at high temperatures, and the rate of refolding would decrease accordingly (Fig. 3). This explanation, although simple and appealing, is not that currently accepted, however. Instead, the temperature-dependence of the rate of refolding is held to be a consequence of the difference in heat capacity between the unfolded and folded states, which must be reflected in non-linear Arrhenius plots of the rates of unfolding or refolding, or both (section 4.D).

3. The pre-folded state. The pre-folded state is the unfolded protein under refolding conditions, prior to the rate-limiting step and complete refolding. Interconversions of the pre-folded state are observed to be rapid relative to the rate of refolding, so the thermodynamic stabilities of its non-random conformations are more relevant for folding than are their rates of formation from the fully unfolded protein.

It is often claimed that secondary structure must be formed before the tertiary structure; this statement is undoubtedly true, but not necessarily relevant. An α-helix forms in any unfolded polypeptide chain on the microsecond time scale [123], but it disappears even more rapidly, for it is usually unstable and the equilibrium constant for its formation is generally less than unity. A more relevant question is at which stage the α-helix becomes stable, but this is likely to be a gradual process, with the equilibrium constant having a variety of values. To what extent must the helix be stable to be significant? There is no meaningful threshold. It might be asked whether a conformation with the helix is on or off the pathway, but this would be impossible to determine kinetically if all the pre-folded state conformations are rapidly interconverted and show similar kinetics of disappearance. These are fundamental difficulties in elucidating the pathway of protein folding in terms of non-covalent interactions [124].

The pre-folded state is intrinsically unstable and populated only transiently. Nevertheless, a variety of evidence indicates that with many proteins it has considerable non-random conformation. For example, all pre-folded states detected by urea-gradient electrophoresis adopted very compact conformations under folding conditions [125]. Evidence is accumulating that the pre-folded state generally is similar to the CI state described in section 2.C [126–128]. The CI state of α-lactalbumin is stable under certain conditions, but not under any known conditions for the homologous hen egg-white lysozyme. Yet, during refolding, both proteins rapidly adopt similar pre-folded conformations that are like the stable CI state of α-lactalbumin [129].

Other aspects of the pre-folded state have been characterized by its susceptibility to proteases. Pre-folded ribonuclease A, with four disulphide bonds intact and at least one incorrect *cis–trans* peptide bond isomer, is protected from cleavage by pepsin near its *C*-terminus [130], whereas another region of the polypeptide chain remains susceptible to cleavage by trypsin [131]. Intriguingly, at least one fragment of dihydrofolate reductase has been shown to inhibit the protein's refolding [132]; the peptide presumably interacts specifically with the pre-folded protein.

One of the potentially most informative probes of the conformational properties of the pre-folded protein uses the susceptibility to exchange with the solvent of the various -NH-groups of the protein. The unfolded protein is isotopically labelled at all exchangeable H atoms in ^2H_2O or ^3H_2O, and refolding is initiated by dilution into ^1H_2O. If the unfolded protein were to remain fully unfolded after dilution, there would be simple competition between refolding and hydrogen exchange of the residual unfolded protein; the fully folded conformation will protect interior groups from subsequent exchange with the solvent. The rates of both hydrogen exchange and refolding are known for the fully unfolded state, so the ^2H or ^3H expected to be retained by the folded protein can be calculated. The observed retention is much greater than this, indicating that hydrogen exchange of the unfolded protein was slower than expected for a fully unfolded polypeptide chain [133]; this suggests that the pre-folded protein did not remain fully unfolded.

The procedure has been extended to determine the locations of the protected -NH-groups, using ^1H-n.m.r. [134,135]. Amide groups with an ^1H atom give an n.m.r. signal, whereas those with an ^2H atom do not. Consequently, the unfolded protein is

initially in ^2H_2O and is transferred to ^1H_2O at acidic pH, where folding occurs but hydrogen exchange is intrinsically slow. At different times, exchange is permitted for a brief period by exposure to high pH. Folding is subsequently allowed to proceed to completion, to protect any buried ^2H atoms that have not exchanged. The fraction of ^1H isotope present at each position in the final refolded protein is determined from two-dimensional n.m.r. spectra.

In assessing these studies, several aspects of the technique must be kept in mind. (1) The only -NH- groups that can be observed are those that are in the interior of the fully folded protein, protected from exchange. Any -NH- group on the surface of the folded protein that might be protected in the pre-folded protein, which would necessarily be by a non-native conformation, could not be detected by the technique. (2) The -NH- groups that are protected from exchange by the folded conformation sufficiently to be observed are usually those involved in secondary structure in the native conformation. (3) The group to which any particular -NH- group is hydrogen bonded in the intermediate cannot be determined from the data. Whether protection was due to α-helix or β-sheet formation might be inferred from the pattern of labelling throughout the polypeptide chain, but only if a single conformation is present in all molecules of the pre-folded state. (4) Incomplete protection of individual -NH- groups may arise from the presence of a mixture of different conformations. Therefore, any patterns of protection might be fortuitous, unless protection is complete in all the molecules. (5) It is virtually impossible to determine directly the kinetic roles of the conformations detected. Unless these considerations are kept in mind, it is very easy to interpret any protected hydrogen atoms as confirming an expectation that native-like secondary structure is present in the pre-folded protein and responsible for the rapid refolding.

Other information pertinent to understanding the pre-folded state comes from protein fragments. Any non-random conformation in such fragments is also likely to be present in the pre-folded protein, unless the other parts of the intact protein actively interfere with it. The occurrence of non-random conformation in protein fragments has only recently been recognized [136]. Previously, short peptides were thought to be unstructured in water, and proteins missing only a few residues from one end were thought to approximate random coils. Although proteins lacking residues from the *C*-terminus are often unfolded [137,138], at least some are far from random coils [139]. The helical tendency of some amino acid sequences [140] is greater than predicted from the classical Zimm–Bragg parameters for the helix-coil transition measured with the host-guest techniques [141]. It is now clear that the Zimm–Bragg formalism is not valid for polypeptides of mixed sequence, due to position-dependent effects and interactions between side-chains [8,142,143]. The intrinsic β-strand tendency is even less certain, for there is no adequate model system for studying β-structure formation experimentally [144].

D. The transition state for folding

The transition state in protein folding is that species along the reaction pathway with the highest free energy, which is encountered in the rate-limiting step (Fig. 6). As with any reaction, its occurrence is hypothetical, and it cannot be characterized directly. Transition states can be characterized only by measuring the effect on the rates of unfolding and refolding of varying either the conditions or the protein. The rate constant is inversely proportional to the relative free energy of the transition state. Under the same conditions, the same transition state should be encountered in both directions of the reaction, i.e., the ratio of the rate constants for unfolding and refolding should be the same as the measured equilibrium constant. That this is usually the case with the relatively simple observed kinetics of protein unfolding and refolding (Fig. 5) suggests that the transition state is a useful concept for a complex reaction like protein folding.

The kinetics of unfolding and refolding usually observed (sections 4.B and 4.C) suggest that the transition state is much closer to the fully folded state than to the unfolded state (Fig. 6*c*). Very substantial conformational changes often precede the rate limiting step in refolding (section 4.C), whereas there is little partial unfolding prior to complete unfolding (Section 4.B). Although the kinetics of unfolding and refolding are measured under identical conditions only within the unfolding transition region, these observations are consistent with the proposal that the transition state for folding is a distorted high-energy form of the native conformation [125]. Presumably because of the co-operativity of the fully folded state, the free energy barrier to unfolding is also the high barrier overall for refolding. This contrasts with nucleation-rapid growth models of folding [145], where the overall transition state would involve a nucleation event in the unfolded conformations (Fig. 6*a*).

The transition state for folding has been most thoroughly analysed in the case of hen egg-white lysozyme [146–148]. Kinetic analysis of the refolding of this protein was relatively simple, because only 10% of the unfolded molecules refold slowly, presumably because there are only two prolyl peptide bonds in its folded conformation, and they are *trans*. The rates of the major, presumably direct, folding process were measured as a function of temperature and denaturant concentrations to characterize the transition state. The majority of the observations indicate that the transition state for lysozyme folding is very similar to the native conformation, although distorted and of high free energy. (1) The linearity of the Arrhenius plot of unfolding rates (fig. 5) suggests that the transition state is not changing and that it has the same heat capacity as the folded state. The non-linear Arrhenius plot for rates of refolding indicate that the heat capacity of the unfolded state is considerably greater than that of the transition and fully folded states. If the heat capacity is reflecting the exposure of non-polar groups to water, the transition state must be very similar to the fully folded protein in this respect. (2) The activation enthalpy for unfolding was found to be independent of temperature and denaturants, and hydrophilic denaturants did not increase the rate of unfolding; both suggest that the transition state excludes water from its interior and is nearly as compact as N. (3) Covalently cross-linking residues Glu-35 and Trp-108 in the folded state altered only the rate of refolding, indicating that these groups were in comparable proximity in the transition state and in N. (4) In contrast, substrate analogues affected primarily the rate of unfolding, indicating that the transition state is distorted sufficiently not to bind ligands specifically.

Qualitatively similar conclusions have been reached about the transition states for folding of other proteins. That in T4 lysozyme folding has 75–80% of the heat capacity of the native protein [149], and its enthalpy is actually greater than that of the fully folded or unfolded states. The transition state in folding of staphylococcal nuclease is similar to the native conformation in that it also does not favour either the *cis* or *trans* isomers of peptide bond 117 [150]. The transition state in unfolding of chymotrypsinogen A is so similar to the fully folded protein that Lumry & Biltonen [151] considered the rate-limiting step in unfolding to be a 'subtle conformational change'.

Altering the protein at specific sites and measuring the effect on the relative stability of N and on the rates of unfolding and refolding has the potential for characterizing the folding transition state in greater detail [152]. Mutations throughout the

protein are generally observed to affect primarily the rate of unfolding, indicating that they alter primarily the energy of the fully folded state, not of the unfolded or transition states. This suggests that the transition state has lost most, if not all, of the co-operativity of the fully folded state, presumably by having lost its close-packing. High pressures are usually observed to slow up folding reactions dramatically [153], indicating that the transition state has an expanded volume relative to both U and N, which barely differ in this respect.

Most of the protein molecule appears to be perturbed in the transition state, although to varying extents. In contrast, Kuwajima *et al.* [154] have proposed that part of the α-lactalbumin molecule is native-like in the transition state and the remainder somewhat unfolded, because this transition state binds Ca^{2+} with an affinity only 10-fold lower than native protein and because the rate of unfolding is increased by denaturants. The increase in the rate of unfolding by denaturants is interpreted as reflecting a greater exposure of non-polar surface in the transition state, but denaturants are known by crystallography to bind directly to the folded states of proteins [81], and might be even more likely to do so in the course of unfolding. Also, the close-packing of the folded interior of a protein might make it more sensitive energetically to weakening of the hydrophobic interaction than the less co-operative transition state.

Current interest in characterizing the transition state for folding suggests that most of these uncertainties should soon be resolved.

5. ELUCIDATING FOLDING PATHWAYS

Elucidating the mechanism of protein folding requires characterization of the initial, final, and intermediate conformational states, plus determination of the steps by which they are interconverted. Elucidating the kinetic roles of the various states requires some means of control over the rates and equilibria of the various steps, which might also make it possible to ensure that normally very unstable intermediates accumulate to substantial levels, at least transiently. Ideally, the unstable intermediates would be trapped in a stable form.

A. Trapping folding intermediates with disulphide bonds

The ideal situation would be to control the rates of formation and breakage of hydrogen bonds, since every protein structure includes them. During folding, molecules with 1, 2, 3,... intramolecular hydrogen bonds might accumulate kinetically; if they could be trapped and identified, a pathway could be defined in terms of hydrogen bonding. It is unfortunately not possible to trap hydrogen bonds, but disulphide bonds can be trapped, due to the reduction–oxidation nature of the covalent disulphide interaction between thiol groups [9,87,155]. Protein species with different numbers of disulphide bonds can be trapped and separated, and their disulphide bonds identified (Fig. 7).

The kinetic roles of the intermediates can be determined relatively unambiguously due to the ability to control the kinetics and thermodynamics of the disulphide interaction. Under the appropriate conditions, the disulphide interaction can be very dynamic, with disulphides being formed, broken and rearranged on time scales as short as at least 10^{-5} s. In this case, the disulphide interaction is similar in many ways to hydrogen bonding [9]. The rates of the intramolecular steps in disulphide formation reflect the protein conformational transitions involved. The approach is only useful with proteins that unfold when their disulphides are broken; unfolding and refolding can then be controlled by varying just the intrinsic disulphide stability. There is no need to use denaturants, and the strengths of all other types of interactions that stabilize proteins can be kept constant.

Although only the disulphide bonds are trapped, the conformations that direct the disulphide bond formation are effectively trapped also (Fig. 8). It is a thermodynamic requirement that whatever conformation stabilized a particular disulphide bond must be stabilized to the very same extent by the presence of that disulphide. Therefore, the conformational basis of folding should be evident from the conformations of the trapped intermediates [156–161], with the proviso that the conformations are not affected by the trapping procedure.

B. Disulphide folding pathway of BPTI

The most detailed and informative folding pathway elucidated thus far is that of bovine pancreatic trypsin inhibitor (BPTI)

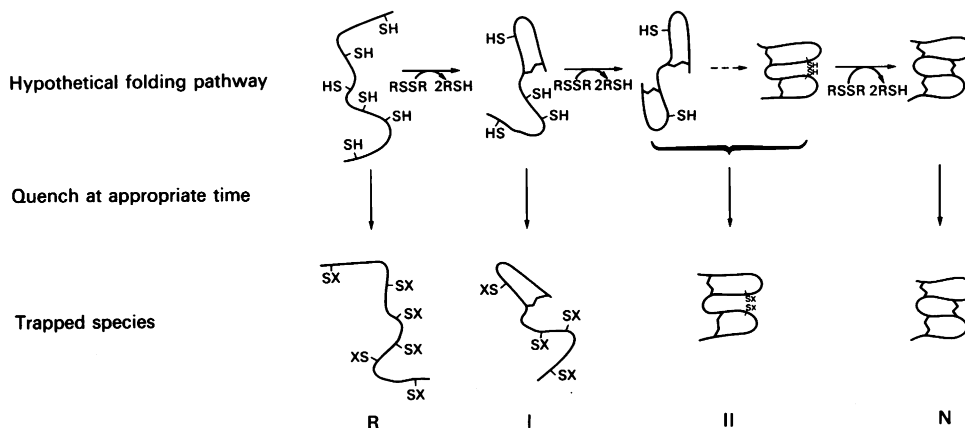


Fig. 7. Rationale of the experimental approach to using disulphide bonds to trap folding intermediates in a stable form

A hypothetical folding pathway of a protein with six cysteine residues is illustrated. Disulphides are generated in the protein by thiol–disulphide interchange with a disulphide reagent, RSSR. A kinetically significant conformational change is illustrated in the two-disulphide intermediate; other conformational changes are assumed to be rapid relative to the rates of disulphide bond formation, breakage and rearrangement. Changes of disulphides can be very rapid and dynamic under the appropriate folding conditions, but can be quenched at any instant of time by rapidly blocking all the free thiol groups with an irreversible reagent, indicated here to add a moiety X. The trapped species are stable indefinitely under the appropriate conditions; they may be separated on the basis of their different conformations, physical properties and numbers of X moieties. The two-disulphide intermediate will be trapped in whatever conformation is the more stable.

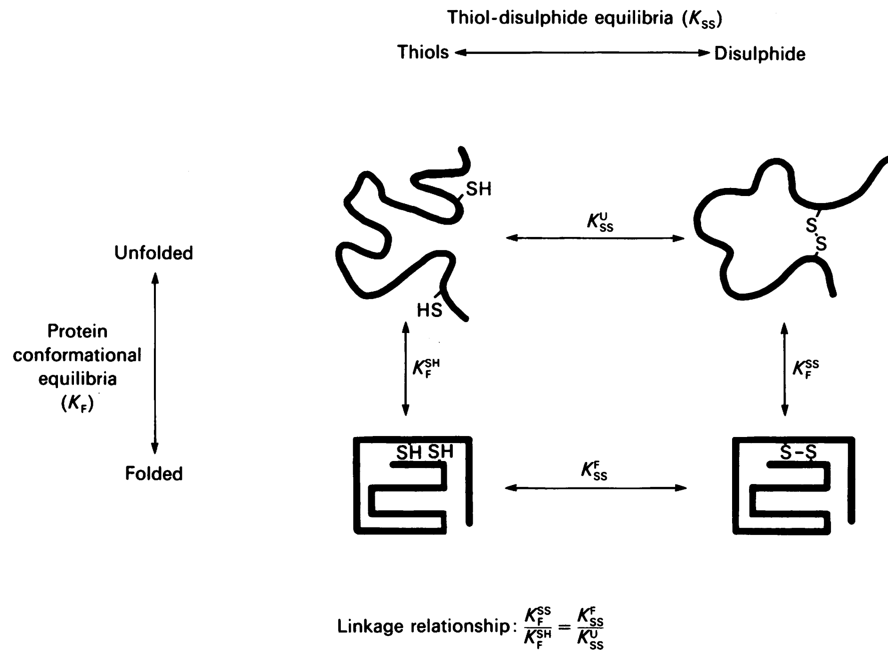


Fig. 8. Protein disulphide bond and conformational stability are linked functions

A protein with two cysteine residues that can form a disulphide bond is illustrated in either the unfolded or folded conformations. The indicated equilibrium constants represent the stabilities of the disulphide bonds, K_{SS}, and of the folded conformation, K_F. The linkage relationship states that whatever effect the folded conformation has on the stability of the disulphide bond, the disulphide bond must have the same effect on the stability of the folded conformation. Comparable linkage relationships pertain to all interactions within the folded conformation, not just disulphide bonds.

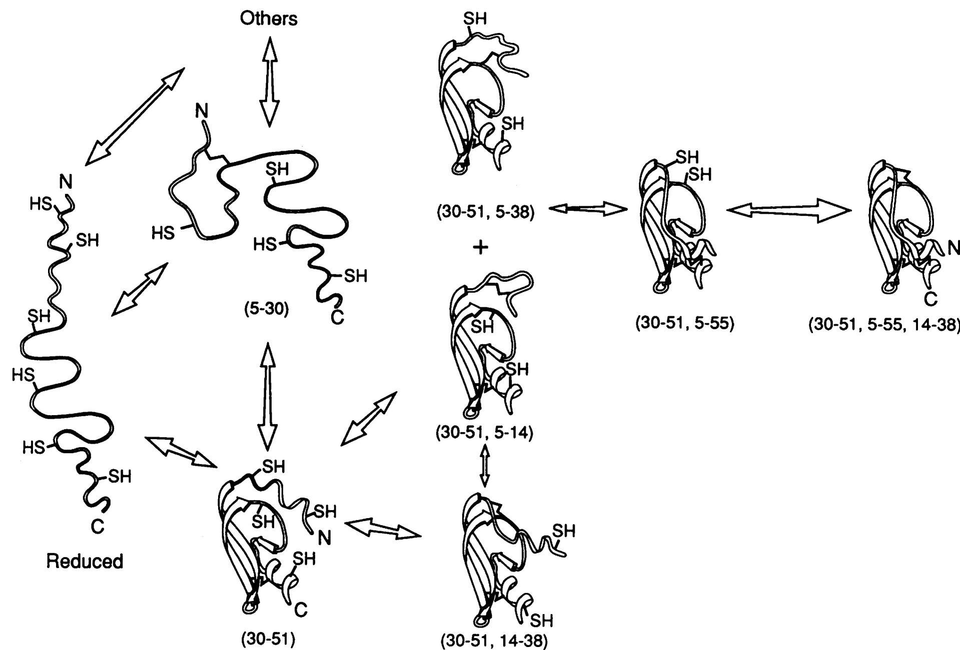


Fig. 9. The disulphide folding pathway of BPTI

The polypeptide backbone of the protein is depicted by a smooth open line when its conformation is not regular or well-defined, by arrows for β-strand conformation, and by a coil for α-helix. The tentative, approximate conformations of the intermediates are based on that of the fully-folded conformation, as drawn by Jane Richardson. The positions of the six cysteine residues are depicted, and the intermediates are designated by the disulphides (solid cross-links) they contain. The relative rates of the intramolecular steps are indicated semi-quantitatively by the thickness of the appropriate arrowhead; the wider the arrowhead, the greater the rate in that direction. The fully reduced protein, R, is unfolded; consequently, formation of the initial disulphide bonds is nearly random. Only the two predominant one-disulphide intermediates are depicted. They are in rapid equilibrium with each other and with the other one-disulphide intermediates. Three different second disulphides, 14–38, 5–14, and 5–38, are formed readily in intermediate (30–51), 10^5-fold more rapidly than is disulphide 5–55. These three second disulphides are rearranged intramolecularly to the native-like intermediate (30–51, 5–55), which readily forms the 14–38 disulphide bond to complete disulphide formation and refolding. The ' + ' between intermediates (30–51, 5–14) and (30–51, 5–38) indicates that they have comparable kinetic roles. A quasi-native (5–55, 14–38) intermediate that is formed directly from minor intermediate (5–55) is not included in this diagram, as it is not on the productive pathway. Unfolding and disulphide breakage occur by the reverse of this pathway.

(Fig. 9). It will be shown here to have all the properties observed for the folding of proteins not involving disulphides.

The three disulphide bonds of native BPTI (between cysteines 5 and 55, 14 and 38, and 30 and 51) are required for stability of the folded conformation. The reduced protein is normally very unfolded, even under physiological conditions, and native-like elements of conformation are barely detectable [162,163]. Whatever the predominant conformations of the reduced protein, they do not lead to productive folding [162]. Even under the most productive folding conditions, formation of the first disulphide bond involves all six cysteine residues in at least approximately random pairings, and the rate is close to that expected for a random coil and is the same as in the presence of 8 M-urea, where random one-disulphide intermediates are generated.

In contrast, the one-disulphide intermediates that actually accumulate under folding conditions are far from random. The intermediate [designated as (30–51)] with the native-like disulphide between Cys-30 and Cys-51 accounts normally for 60% of the one-disulphide molecules, the non-native (5–30) comprises another 30%, while the other 13 possible disulphides comprise the remaining 10%. The explanation for this apparent paradox is that whatever disulphide is formed initially is rapidly rearranged intramolecularly, by thiol–disulphide interchange, and the intermediates that accumulate reflect the equilibrium mixture of one-disulphide species. Those that accumulate to the greatest extent are those with the lowest free energies, as a result of their favourable conformational properties, which in turn depends upon both the conditions and the sequence of the protein.

The preferential accumulation of the most stable intermediate (30–51) has important kinetic consequences, for the productive pathway for refolding leads from this intermediate, and all further productive intermediates retain the 30–51 disulphide bond. The rapid equilibration of the one-disulphide species demonstrates how unfolded proteins can equilibrate rapidly prior to refolding, how all unfolded molecules can follow the same pathway, and why the initial state of the unfolded protein is not important in determining the rate or pathway of folding.

The conformational properties of intermediate (30–51) that account for its preferential stability and its role in further refolding are of greatest importance for understanding this folding mechanism. In native BPTI, the 30–51 disulphide links the major α-helix of the protein to the β sheet, raising the possibility that the stability of the (30–51) intermediate arises from an interaction between these two elements of secondary structure [164]. Evidence for the presence of the secondary structure was obtained only with a homologous protein [161,165], because the aromatic side-chains of BPTI contribute to its far-u.v. c.d. spectrum [159,166], until n.m.r. analysis detected the presence of the β-sheet [160]. That both elements of secondary structure are probably present comes from the work of Oas & Kim [167], who showed that two synthetic peptides of 16 and 14 residues, respectively spanning the sequence around Cys-30 and Cys-51, adopt these conformations when linked by the disulphide. This simple system appears to be a remarkably good model for the (30–51) intermediate, which would seem to have about half the polypeptide chain in a relatively fixed conformation, with the remainder flexible. This evidence suggests that the conformation stabilizing this crucial early intermediate in refolding arises from the interaction between the two major elements of secondary structure of the protein, which also involves much of the hydrophobic interior of the protein. The conformational interactions that direct the folding pathway are also being investigated using site-directed mutagenesis [168].

The conformation present in (30–51) is not highly populated in the individual model peptides, in the absence of the disulphide bond, and is probably present in reduced BPTI in no more than 0.1% of the molecules [162]. The predominant stability of (30–51) relative to the other one-disulphide intermediates is not a result of the pre-existence of its favourable conformation in the reduced protein, but is due to its reciprocal stabilization by the disulphide bond (Fig. 8).

Of the three disulphides in native BPTI, the only one well-populated at the single-disulphide state is 30–51, even though 5–55 is more stable in the fully folded conformation [91], which demonstrates that the most stable parts of a fully folded protein are not necessarily those that are formed initially in folding. The corresponding one-disulphide intermediate, (5–55), is present as only about 3% of the one-disulphide intermediates. The most likely explanation for the high stability of the 5–55 disulphide in native BPTI is that it results from simultaneous interactions between residues in three different regions of the polypeptide chain: those around each of Cys-5 and Cys-55, plus the β-sheet. The simultaneous presence of all three regions of the polypeptide chain would be entropically unfavourable at an early stage of folding; any necessity to bury the cysteine 30 and 51 thiols [169] would also de-stabilize the (5–55) single-disulphide intermediate. The other native disulphide, 14–38, is less stable than 30–51 in native BPTI and is not present at a detectable level in the one-disulphide intermediates. Why intermediate (5–30), with a non-native disulphide, is the second most favoured one-disulphide intermediate is not apparent.

The presence of native-like β-sheet and α-helix in intermediate (30–51), with the remainder of the polypeptide chain disordered, is consistent with the tendency of the second disulphide bond to be formed between any pair of cysteine residues 5, 14 and 38, at a rate comparable to forming the first disulphide. That the disulphide bond between Cys-14 and Cys-38 is formed at such a relatively low rate indicates that the native-like conformation in (30–51) does not extend to both Cys-14 and Cys-38; otherwise, this disulphide should be formed at least 200-fold more rapidly. Cys-55 of intermediate (30–51) does not readily form a disulphide with any of the other three Cys residues. It is often claimed that this could be due to inaccessibility of the Cys-55 thiol group [169,170], but this thiol is observed experimentally to be accessible and normally reactive [171,172]. The most likely reason why Cys-55 does not participate in disulphide formation is that the non-random conformation of (30–51) prevents Cys-55 from encountering Cys-5, -14 and -38. Although formation of a disulphide with Cys-5 would produce the native-like two-disulphide species (30–51, 5–55), this step probably is so slow because it involves traversing the high energy barrier that separates the more unfolded species from the native conformation. Probably for the same reason, intermediate (30–51, 14–38) does not readily complete refolding by forming directly the 5–55 disulphide.

The rate-limiting step in BPTI refolding occurs just before reaching the stable native conformation, and the pathway of unfolding indicates that the overall transition state is a distorted form of the folded conformation. The most favourable pathway energetically into and out of the native conformation of BPTI is by intramolecular rearrangement of the non-native second disulphides of intermediates (30–51, 5–14) and (30–51, 5–38) to the native-like (30–51, 5–55). That the energetically most favourable pathway into and out of the native conformation of BPTI is via these disulphide rearrangements is believed to reflect the exceptionally high stability of the native conformation of BPTI, distortion of which is also exceptionally difficult. The disulphide rearrangement pathway is not the most favourable energetically with less stable homologues of BPTI [173], which may be more typical of small proteins. With these less stable proteins, formation of the 5–55 disulphide in intermediate (30–51) is the

energetically preferred, although slow, rate-limiting step in refolding. Consequently, the disulphides of these proteins are formed in a seemingly simple, sequential manner, first 30–51, then 5–55, finally 14–38, but the slowness of the rate-limiting step indicates that forming the 5–55 disulphide probably still involves a distortion of the native-like conformation. This might be analogous to the disulphide rearrangement pathway, although on a smaller scale, but in this case there are no Cys residues in appropriate positions to detect these distortions by disulphides.

Once the transition state has been overcome, the native-like intermediate (30–51, 5–55) results. The two remaining free Cys residues are held in proximity by the conformation and are on the surface of the molecule, where distortions are not so energetically unfavourable, so the 14–38 disulphide bond is formed rapidly.

Unfolding and disulphide breakage of these proteins occur by the reverse of this process merely by destabilizing the disulphide interaction. The height of the free energy barrier to unfolding is observed to be inversely proportional to the stability of the folded state. The disulphide rearrangements of BPTI demonstrate vividly, if to a somewhat exaggerated extent, the importance of the high free-energy barrier of the distorted native-like conformation (section 4.D).

The disulphide intermediates are less stable than either the fully reduced or fully folded states under all known conditions, and have less stable folded conformations than N, so the BPTI disulphide folding transition demonstrates the usual cooperativity of folding (section 3.B). All but the rate-limiting steps are readily reversible, so the initial one- and two-disulphide intermediates rapidly equilibrate with the fully reduced protein prior to refolding, and (30–51, 5–55) rapidly equilibrates with native protein prior to complete unfolding.

Similar energetics have been observed with the other proteins that have been examined in this way, especially ribonuclease T1 [174] and ribonuclease A [117,175]. Intramolecular rearrangements of disulphide bonds appear to be important and rate-limiting in the disulphide folding of these proteins. It is not surprising, therefore, that proteins capable of catalysing the disulphide rearrangements exist [18]. The best-characterized protein-disulphide isomerase catalyzes all the steps in the BPTI folding pathway that involve both substantial conformational changes and disulphide bond formation, breakage or rearrangement [176].

C. A current view of protein folding pathways

With current knowledge of protein folding, there is no conclusive evidence to indicate that the folding pathway of BPTI and its homologues is not typical of small single-domain proteins in general. If so, it is possible to extrapolate from one to the other by equating alteration of the stability of the disulphide interaction with, for example, changing the denaturant concentration or temperature to vary the folding conditions not involving disulphides.

Upon placing an initially unfolded protein into conditions favouring folding, most proteins appear to adopt, rapidly and reversibly, a limited number of non-random conformations, the pre-folded state (Fig. 6*c*). The equivalent state in BPTI would be the mixture of the initial one- and two-disulphide intermediates (Fig. 9). If the pre-folded state is like the CI state (section 2.C), this would imply that the one- and two-disulphide intermediates of the BPTI are collectively analogous to the CI state of other proteins. This hypothesis is supported by the recent findings that the CI state of α-lactalbumin has some elements of relatively fixed conformation, plus other more flexible parts of the polypeptide chain [52], similar to the BPTI early intermediates. On the other hand, the BPTI intermediates demonstrate evidence of aromatic side-chains in asymmetric environments [159] that is not typical of the CI states that have been well-characterized (section 2.C). More evidence is needed to test this hypothesis.

The initial acquisition of non-random conformation in an unfolded protein may occur randomly; some of these non-random conformations will be more stable than others, will predominate, and some may be important for acquiring rapidly further non-random conformation. All of these non-random conformations will be only meta-stable, however, so they will also unfold rapidly. Under analogous conditions with BPTI and its homologues, the initial one- and two-disulphide intermediates are made and unfolded some 10^4–10^5 times before folding is completed. Accordingly, much of the time preceding complete folding would be taken up by molecules interconverting rapidly between a few conformations, plus the unfolded state (Fig. 6*c*).

The primary driving force for adopting non-random conformation is most likely to be the hydrophobic effect. A hydrophobic collapse would greatly increase the stability of hydrogen-bonding of the polypeptide backbone, so hydrophobic interactions between the intrinsically most stable elements of secondary structure would not be unexpected. Other things being equal, interactions between groups close in the primary structure should be most stable, for entropic reasons. This could explain why elements of secondary structure adjacent in the primary structure tend to be adjacent also in the final folded conformation [177].

Many of the favourable elements of conformation in the pre-folded state could be similar to those in the native conformation, so some of the molecules will transiently adopt sufficient native-like conformation to approach that of the transition state. They will then rapidly complete refolding, rather than spontaneously unfolding. If the transition state is a high-energy form of the native-like conformation, conformational rearrangements and distortion of native-like elements already present may occur, analogous to, but perhaps less extreme than, the disulphide rearrangements of the BPTI pathway (Fig. 9).

Once through the transition state, the majority of the native conformation, at least the close-packed interior, should be present, and minor adjustments on the surface, analogous to forming the 14–38 disulphide of BPTI, can occur extremely rapidly.

The validity of the above scenario needs to be tested by more thorough studies of the folding pathways of other proteins. The current pace of progress in the field indicates that there should not be long to wait.

REFERENCES

1. Regan, L. & DeGrado, W. F. (1988) Science **241**, 976–978
2. DeGrado, W. F. (1988) Adv. Protein Chem. **39**, 51–124
3. Goraj, K., Renard, A. & Martial, J. A. (1990) Protein Eng. **3**, 259–266
4. Anfinsen, C. B. (1973) Science **181**, 223–230
5. Anfinsen, C. B. & Scheraga, H. A. (1975) Adv. Protein Chem. **29**, 205–300
6. Baldwin, R. L. (1975) Annu. Rev. Biochem. **44**, 453–475
7. Baldwin, R. L. (1986) Trends Biochem. Sci. **11**, 6–9
8. Baldwin, R. L. (1989) Trends Biochem. Sci. **14**, 291–294
9. Creighton, T. E. (1978) Prog. Biophys. Mol. Biol. **33**, 231–297
10. Baldwin, R. L. & Creighton, T. E. (1980) in Protein Folding (Jaenicke, R., ed.), pp. 217–260, Elsevier/North-Holland, Amsterdam
11. Kim, P. S. & Baldwin, R. L. (1982) Annu. Rev. Biochem. **51**, 459–489
12. Kim, P. S. & Baldwin, R. L. (1990) Annu. Rev. Biochem., in the press
13. Jaenicke, R. (1984) Angew. Chem. Int. Ed. Engl. **23**, 395–413
14. Jaenicke, R. (1987) Prog. Biophys. Mol. Biol. **49**, 117–237
15. Pelham, H. (1988) Nature (London) **332**, 776–777
16. Randall, L. L. & Hardy, S. J. S. (1989) Science **243**, 1156–1159

17. Hartl, F.-U. & Neupert, W. (1990) Science **247**, 930–938
18. Freedman, R. B. (1989) Cell **57**, 1069–1072
19. Fischer, G. & Schmid, F. X. (1990) Biochemistry **29**, 2205–2212
20. Campbell, I. D., Dobson, C. M. & Williams, R. J. P. (1985) Biochem. J. **231**, 1–10
21. Wüthrich, K. (1986) NMR of Proteins and Nucleic Acids, Wiley, New York
22. Wüthrich, K. (1989) Acc. Chem. Res. **22**, 36–44
23. Wüthrich, K. (1989) Science **243**, 45–50
24. Clore, G. M. & Gronenborn, A. M. (1989) Crit. Rev. Biochem. **24**, 479–564
25. Schultz, G. E. & Schirmer, R. H. (1979) Principles of Protein Structure, Springer-Verlag, New York
26. Richardson, J. S. (1981) Adv. Protein Chem. **50**, 497–532
27. Creighton, T. E. (1983) Proteins: Structures and Molecular Properties, W. H. Freeman, New York
28. Chothia, C. (1984) Annu. Rev. Biochem. **53**, 537–572
29. Bajaj, M. & Blundell, T. (1984) Annu. Rev. Biophys. Bioeng. **13**, 453–492
30. Chothia, C. & Lesk, A. M. (1986) EMBO J. **5**, 823–826
31. Perutz, M. F., Kendrew, J. C. & Watson, H. C. (1965) J. Mol. Biol. **13**, 669–678
32. Lim, W. A. & Sauer, R. T. (1989) Nature (London) **339**, 31–36
33. Kellis, J. T., Jr., Nyberg, K. & Fersht, A. R. (1989) Biochemistry **28**, 4914–4922
34. Bowie, J. U., Reidhaar-Olson, J. F., Lim, W. A. & Sauer, R. T. (1990) Science **247**, 1306–1310
35. McCammon, J. A. & Harvey, S. C. (1987) Dynamics of Proteins and Nucleic Acids, Cambridge University Press, London
36. Smith, J. L., Hendrickson, W. A., Honzatko, R. B. & Sheriff, S. (1986) Biochemistry **25**, 5018–5027
37. Wagner, G. (1983) Q. Rev. Biophys. **16**, 1–57
38. Chothia, C., Lesk, A. M., Dodson, G. & Hodgkin, D. C. (1983) Nature (London) **302**, 500–505
39. Löbermann, H., Tokuoka, R., Deisenhofer, J. & Huber, R. (1984) J. Mol. Biol. **177**, 531–556
40. Perutz, M. F. (1989) Q. Rev. Biophys. **22**, 139–236
41. Levinthal, C. (1968) J. Chim. Phys. **65**, 44–45
42. Tanford, C. (1968) Adv. Protein Chem. **23**, 121–282
43. Flory, P. J. (1969) Statistical Mechanics of Chain Molecules, Wiley, New York
44. Miller, W. G., Brant, D. A. & Flory, P. J. (1967) J. Mol. Biol. **23**, 67–80
45. Dobson, C. M., Evans, P. A. & Williamson, K. L. (1984) FEBS Lett. **168**, 331–334
46. Howarth, O. W. & Lian, L. Y. (1984) Biochemistry **23**, 3515–3521
47. Griko, Yu, V., Privalov, P. L., Venyaminov, S., Yu. & Kutyshenko, V. P. (1988) J. Mol. Biol. **202**, 127–138
48. Privalov, P. L., Tiktopulo, E. I., Venyaminov, S. Yu., Griko, Yu. V., Makhatadze, G. I. & Khechinashvili, N. N. (1989) J. Mol. Biol. **205**, 737–750
49. Privalov, P. L. (1979) Adv. Protein Chem. **33**, 167–241
50. Ptitsyn, O. B. (1987) J. Protein Chem. **6**, 273–293
51. Kuwajima, K. (1989) Proteins: Struct. Funct. Genet. **6**, 87–103
52. Baum, J., Dobson, C. M., Evans, P. A. & Hanley, C. (1989) Biochemistry **28**, 7–13
53. Tweedy, N. B., Hurle, M. R., Chrunyk, B. A. & Matthews, C. R. (1990) Biochemistry **29**, 1539–1545
54. Privalov, P. L. (1982) Adv. Protein Chem. **35**, 1–104
55. Brandts, J. F., Ju, C. Q., Lin, L.-N. & Mas, M. T. (1989) Biochemistry **28**, 8588–8596
56. Griko, Yu. V., Venyaminov, S. Yu. & Privalov, P. L. (1989) FEBS Lett. **244**, 276–278
57. Schellman, J. A. (1987) Annu. Rev. Biophys. Biophys. Chem. **16**, 115–137
58. Ahmad, F. & Bigelow, C. C. (1986) Biopolymers, **25**, 1623–1633
59. Pace, C. N., Laurents, D. V. & Thomson, J. A. (1990) Biochemistry **29**, 2564–2572
60. Privalov, P. L. (1989) Annu. Rev. Biophys. Biophys. Chem. **18**, 47–69
61. Sturtevant, J. M. (1977) Proc. Natl. Acad. Sci. U.S.A. **74**, 2236–2240
62. Spolar, R. S., Ha, J.-H. & Record, M. T., Jr. (1989) Proc. Natl. Acad. Sci. U.S.A. **86**, 8382–8385
63. Becktel, W. J. & Schellman, J. A. (1987) Biopolymers **26**, 1859–1877
64. Baldwin, R. L. (1986) Proc. Natl. Acad. Sci. U.S.A. **83**, 8069–8072
65. Privalov, P. L., Griko, Yu. V., Venyaminov, S. Yu. & Kutyshenko, V. P. (1986) J. Mol. Biol. **190**, 487–498
66. Chen, B.-I. & Schellman, J. A. (1989) Biochemistry **28**, 685–691
67. Kauzmann, W. (1959) Adv. Protein Chem. **14**, 1–63
68. Privalov, P. L. & Gill, S. J. (1988) Adv. Protein Chem. **39**, 191–234
69. Murphy, K. P., Privalov, P. L. & Gill, S. J. (1990) Science **247**, 559–561
70. Muller, N. (1990) Acc. Chem. Res. **23**, 23–28
71. Creighton, T. E. (1983) Biopolymers **22**, 49–58
72. Page, M. I. & Jencks, W. P. (1971) Proc. Natl. Acad. Sci. U.S.A. **68**, 1678–1683
73. Baker, E. N. & Hubbard, R. E. (1984) Prog. Biophys. Mol. Biol. **44**, 97–179
74. Bello, J. (1977) J. Theor. Biol. **68**, 139–142
75. Alber, T., Dao-Pin, S., Wilson, K., Wozniak, J. A., Cook, S. P. & Matthews, B. W. (1987) Nature (London) **330**, 41–46
76. Serrano, L. & Fersht, A. R. (1989) Nature (London) **342**, 296–299
77. Anderson, D. E., Becktel, W. J. & Dahlquist, F. W. (1990) Biochemistry **29**, 2403–2408
78. Nall, B. T. & Baldwin, R. L. (1977) Biochemistry **16**, 3572–3576
79. Schwarz, F. P. (1988) Biochemistry **27**, 8429–8436
80. Liu, W. & Tsou, C.-L. (1987) Biochim. Biophys. Acta **916**, 455–464
81. Hibbard, L. S. & Tulinsky, A. (1978) Biochemistry **17**, 5460–5468
82. Puett, D. (1973) J. Biol. Chem. **248**, 4623–4634
83. Matthews, B. W. (1987) Biochemistry **26**, 6885–6888
84. Fersht, A. R. (1987) Biochemistry **26**, 8031–8037
85. Goldenberg, D. P. (1988) Annu. Rev. Biophys. Biophys. Chem. **17**, 481–507
86. Shortle, D. (1989) J. Biol. Chem. **264**, 5315–5318
87. Creighton, T. E. (1986) Methods Enzymol. **131**, 83–106
88. Kellis, J. T., Jr., Nyberg, K., Sali, D. & Fersht, A. R. (1988) Nature (London) **333**, 784–786
89. Matsumara, M., Becktel, W. J. & Matthews, B. W. (1988) Nature (London) **334**, 406–410
90. Garvey, E. P. & Matthews, C. R. (1989) Biochemistry **28**, 2083–2093
91. Creighton, T. E. & Goldenberg, D. P. (1984) J. Mol. Biol. **179**, 497–526
92. Alber, T., Dao-pin, S., Nye, J. A., Muchmore, D. C. & Matthews, B. W. (1987) Biochemistry **26**, 3754–3758
93. Ikemura, H., Takagi, H. & Inouye, M. (1987) J. Biol. Chem. **262**, 7859–7864
94. Silen, J. L., Frank, D., Fujishige, A., Bone, R. & Agard, D. A. (1989) J. Bacteriol. **171**, 1320–1325
95. West, S. M., Kelly, S. M. & Price, N. C. (1990) Biochim. Biophys. Acta **1037**, 332–336
96. West, S. M. & Price, N. C. (1990) Biochem. J. **265**, 45–50
97. Villafane, R. & King, J. (1988) J. Mol. Biol. **204**, 607–619
98. Goldenberg, D. P. & Creighton, T. E. (1983) J. Mol. Biol. **165**, 407–413
99. Luger, K., Hommel, U., Herold, M., Hofsteenge, J. & Kirschner, K. (1989) Science **243**, 206–210
100. Harrison, S. C. & Durbin, R. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**, 4028–4030
101. Brandts, J. F., Halvorson, H. R. & Brennan, M. (1975) Biochemistry **14**, 4953–4963
102. Evans, P. A., Dobson, C. M., Kautz, R. A., Hatfull, G. & Fox, R. O. (1987) Nature (London) **329**, 266–268
103. Chazin, W. J., Kördel, J., Drakenberg, T., Thulin, E., Brodin, P., Grundström, T. & Forsén, S. (1989) Proc. Natl. Acad. Sci. U.S.A. **86**, 2195–2198
104. Creighton, T. E. (1978) J. Mol. Biol. **125**, 401–406
105. Schmid, F. X. (1986) FEBS Lett. **198**, 217–220
106. Schmid, F. X., Grafl, R., Wrba, A. & Beintema, J. J. (1986) Proc. Natl. Acad. Sci. U.S.A. **83**, 872–876
107. Lin, L.-N. & Brandts, J. F. (1988) Biochemistry **27**, 9037–9042
108. Ramdas, L. & Nall, B. T. (1986) Biochemistry **25**, 6959–6964
109. Kelley, R. F. & Richards, F. M. (1987) Biochemistry **26**, 6765–6774
110. Kiefhaber, T., Grunert, H.-P., Hahn, U. & Schmid, F. X. (1990) Biochemistry, in the press
111. Lang, K., Schmid, F. X. & Fischer, G. (1987) Nature (London) **329**, 268–270
112. Fischer, G., Wittman-Liebold, B., Lang, K., Kiefhaber, T. & Schmid, F. X. (1989) Nature (London) **337**, 476–478
113. Takahashi, N., Hayano, T. & Suzuki, M. (1989) Nature (London) **337**, 473–475
114. Siekierka, J. J., Hung, S. H. Y., Poe, M., Lin, C. S. & Sigal, N. H. (1989) Nature (London) **341**, 755–757
115. Harding, M. W., Galat, A., Uehling, D. E. & Schreiber, S. L. (1989) Nature (London) **341**, 758–760

116. Brandts, J. F., Brennan, M. & Lin, L.-N. (1977) Proc. Natl. Acad. Sci. U.S.A. **74**, 4178–4181
117. Creighton, T. E. (1988) Proc. Natl. Acad. Sci. U.S.A. **85**, 5082–5086
118. Garel, J.-R., Nall, B. & Baldwin, R. L. (1976) Proc. Natl. Acad. Sci. U.S.A. **73**, 1853–1857
119. Kato, S., Okamura, M., Shimamoto, N. & Utiyama, H. (1981) Biochemistry **20**, 1080–1085
120. Denton, J. B., Konishi, Y. & Scheraga, H. A. (1982) Biochemistry **21**, 5155–5163
121. Lynn, R. M., Konishi, Y. & Scheraga, H. A. (1984) Biochemistry **23**, 2470–2477
122. Lang, K. & Schmid, F. X. (1990) J. Mol. Biol. **212**, 185–196
123. Gruenewald, B., Nicola, C. U., Lustig, A., Schwarz, G. & Klump, H. (1979) Biophys. Chem. **9**, 137–147
124. Creighton, T. E. (1977) J. Mol. Biol. **113**, 313–328
125. Creighton, T. E. (1980) J. Mol. Biol. **137**, 61–80
126. Kuwajima, K., Yamaya, H., Miwa, S., Sugai, S. & Magamura, T. (1987) FEBS Lett. **221**, 115–118
127. Semisotnov, G. V., Rodionova, N. A., Kutyshenko, V. P., Ebert, B., Blanck, J. & Ptitsyn, O. B. (1987) FEBS Lett. **224**, 9–13
128. Ptitsyn, O. B., Pain, R. H., Semisotnov, G. V., Zerovnik, E. & Razgulyaev, O. I. (1990) FEBS Lett. **262**, 20–24
129. Ikeguchi, M., Kuwajima, K., Mitani, M. & Sugai, S. (1986) Biochemistry **15**, 6965–6972
130. Schmid, F. & Blaschek, H. (1984) Biochemistry **23**, 2128–2133
131. Lang, K. & Schmid, F. X. (1986) Eur. J. Biochem. **159**, 275–281
132. Hall, J. G. & Frieden, C. (1989) Proc. Natl. Acad. Sci. U.S.A. **86**, 3060–3064
133. Schmid, F. X. & Baldwin, R. L. (1979) J. Mol. Biol. **135**, 199–215
134. Udgaonkar, J. B. & Baldwin, R. L. (1988) Nature (London) **335**, 694–699
135. Röder, H., Elöve, G. A. & Englander, S. W. (1988) Nature (London) **335**, 700–704
136. Wright, P. E., Dyson, H. J. & Lerner, R. A. (1988) Biochemistry **27**, 7167–7175
137. Taniuchi, H. (1970) J. Biol. Chem. **245**, 5459–5468
138. Sachs, D. H., Schechter, A. N., Eastlake, A. & Anfinsen, C. B. (1974) Nature (London) **251**, 242–244
139. Shortle, D. & Meeker, A. K. (1989) Biochemistry **28**, 936–944
140. Marqusee, S., Robbins, V. H. & Baldwin, R. L. (1989) Proc. Natl. Acad. Sci. U.S.A. **86**, 5286–5290
141. Sueki, M., Lee, S., Powers, S. P., Denton, J. B., Konishi, Y. & Scheraga, H. A. (1984) Macromolecules **17**, 148–155
142. Scheraga, H. A. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**, 5585–5587
143. Padmanabhan, S., Marqusee, S., Ridgeway, T., Laue, T. M. & Baldwin, R. L. (1990) Nature (London) **344**, 268–270
144. Mattice, W. L. (1989) Annu. Rev. Biophys. Biophys. Chem. **18**, 93–111
145. Wetlaufer, D. B. (1973) Proc. Natl. Acad. Sci. U.S.A. **70**, 697–701
146. Segawa, S.-I. & Sugihara, M. (1984) Biopolymers **23**, 2473–2488
147. Segawa, S.-I. & Sugihara, M. (1984) Biopolymers **23**, 2489–2498
148. Segawa, S.-I. & Kuma, K. (1986) Biopolymers **25**, 1981–1996
149. Chen, B.-I., Baase, W. A. & Schellman, J. A. (1989) Biochemistry **28**, 691–699
150. Evans, P. A., Kautz, R. A., Fox, R. O. & Dobson, C. M. (1989) Biochemistry **28**, 362–370
151. Lumry, R. & Biltonen, R. (1969) in Structure and Stability of Biological Macromolecules (Timasheff, S. N. & Fasman, G. D., eds.), pp. 65–212, Marcel Dekker, New York
152. Matouschek, A., Kellis, J. T., Jr., Serrano, L. & Fersht, A. R. (1989) Nature (London) **340**, 122–126
153. Hawley, S. A. & Mitchell, R. M. (1975) Biochemistry **14**, 3257–3264
154. Kuwajima, K., Mitani, M. & Sugai, S. (1989) J. Mol. Biol. **206**, 547–561
155. Creighton, T. E. (1988) BioEssays **8**, 57–64
156. Creighton, T. E., Kalef, E. & Arnon, R. (1978) J. Mol. Biol. **123**, 129–147
157. Kosen, P. A., Creighton, T. E. & Blout, E. R. (1980) Biochemistry **19**, 4936–4944
158. Kosen, P. A., Creighton, T. E. & Blout, E. R. (1981) Biochemistry **20**, 5744–5754
159. Kosen, P. A., Creighton, T. E. & Blout, E. R. (1983) Biochemistry **22**, 2433–2440
160. States, D. J., Creighton, T. E., Dobson, C. M. & Karplus, M. (1987) J. Mol. Biol. **195**, 731–739
161. Hollecker, M. & Larcher, D. (1989) Eur. J. Biochem. **179**, 87–94
162. Creighton, T. E. (1988) Biophys. Chem. **31**, 155–162
163. Goodman, E. M. & Kim, P. S. (1989) Biochemistry **28**, 4343–4347
164. Creighton, T. E. (1974) J. Mol. Biol. **87**, 603–624
165. Hollecker, M., Creighton, T. E. & Gabriel, M. (1981) Biochimie **63**, 835–839
166. Manning, M. C. & Woody, R. W. (1989) Biochemistry **28**, 8609–8613
167. Oas, T. G. & Kim, P. (1988) Nature (London) **336**, 42–48
168. Goldenberg, D. P., Frieden, R. W., Haack, J. A. & Morrison, T. B. (1989) Nature (London) **338**, 127–132
169. Staley, J. P. & Kim, P. S. (1990) Nature (London) **344**, 685–688
170. Goto, Y. & Hamaguchi, K. (1981) J. Mol. Biol. **146**, 321–340
171. Creighton, T. E. (1975) J. Mol. Biol. **96**, 777–782
172. Creighton, T. E. (1981) J. Mol. Biol. **151**, 211–213
173. Hollecker, M. & Creighton, T. E. (1983) J. Mol. Biol. **168**, 409–437
174. Pace, C. N. & Creighton, T. E. (1986) J. Mol. Biol. **188**, 477–486
175. Wearne, S. J. & Creighton, T. E. (1988) Proteins **4**, 251–261
176. Creighton, T. E., Hillson, D. A. & Freeman, R. B. (1980) J. Mol. Biol. **142**, 43–62
177. Levitt, M. & Chothia, C. (1976) Nature (London) **261**, 552–558
178. Pfeil, W. & Privalov, P. L. (1976) Biophys. Chem. **4**, 23–50
179. Gold, V. (1979) Pure Appl. Chem. **51**, 1725–1801
180. Yagisawa, S. (1989) Biochem. J. **263**, 985–988