# How do proteins fold?

**Utpal Nath and Jayant B. Udgaonkar**

National Centre for Biological Sciences, TIFR Centre, P.O. Box 1234, Indian Institute of Science Campus, Bangalore 560 012, India

**Understanding the mechanism by which an unfolded polypeptide chain folds to its unique, functional structure is a primary unsolved problem in biochemistry. Fundamental advances towards understanding how proteins fold have come from kinetic studies. Kinetic studies allow the dissection of the folding pathway of a protein into individual steps that are defined by partially-structured folding intermediates. Improvements in both the structural and temporal resolution of physical methods that are used to monitor the folding process, as well as the development of new methodologies, are now making it possible to obtain detailed structural information on protein folding pathways. The protein engineering methodology has been particularly useful in characterizing the structures of folding intermediates as well as the transition state of folding. Several characteristics of protein folding pathways have begun to emerge as general features for the folding of many different proteins. Progress in our understanding of how structure develops during folding is reviewed here.**

PROTEINS fold to highly intricate and specific structures following their synthesis on ribosomes. The folding of nascent polypeptide chains to their native structures is an essential step in decoding genetic information to cellular activities, because only the folded conformation of a protein is functional. In contrast to our knowledge on transcription and translation as initial steps in information transfer in the cell, little is known about how proteins fold. Obtaining an understanding of the principles governing the formation of a fully-folded native protein from its unfolded state forms the basis of 'the protein folding problem'.

The earliest studies on protein folding date back to the 1920s, when Anson and Mirsky reported, for the first time, that the denaturation of many proteins is reversible[1,2]. This observation was extended by others[3,4], but it was not until 1961 that it was convincingly proven that the three-dimensional structure of a protein is dictated solely by its primary sequence[5,6]. In their classic work on bovine pancreatic ribonuclease A (RNase A), Anfinsen and his colleagues showed that reduced, denatured RNase A refolds spontaneously to its native conformation *in vitro* with the restoration of enzymatic activity as well as formation of its four native disulphide bridges, in the absence of any external assistance[7]. Since this landmark experiment of Anfinsen and his colleagues, research on protein folding has been proceeding at an explosive rate in many laboratories worldwide. Many excellent articles and books have reviewed different aspects of the protein folding problem[8–44]. This review will discuss the results of recent experimental studies that utilized kinetic methods, and will place special emphasis on how structure develops.

A review on protein folding remains incomplete without a discussion of the possible practical applications of protein folding studies[11,45]: (i) Virtually any foreign gene can be expressed in a host bacterium or a cell line using recombinant DNA technology. The expression of the recombinant gene often leads to the production of the protein in an inactive insoluble form, known as inclusion bodies[46]. Recovery of the active forms of these recombinant proteins from their inactive forms requires a good practical understanding of protein folding. (ii) Recombinant DNA techniques and modern peptide synthesis techniques permit the synthesis of large polypeptide chains of any desired sequence. Since the function of a protein is dictated by its three-dimensional structure, it is necessary to know, *a priori*, how a specific amino acid sequence will fold once synthesized. Though reports of designed amino acid sequences which adopt intended folds and possess physical properties similar to those of natural proteins are appearing with increasing frequency[40,47–50], designing proteins with novel biological functions still remains a major challenge. (iii) Although recombinant DNA techniques can be used to modify, at will, the amino acid sequence of an existing protein, the ability to do such protein-engineering in a rational manner depends on first gaining an understanding of how changing the sequence of a protein will alter its structure. (iv) Many diseases, including cystic fibrosis, scurvy, scrapie and Alzheimer's disease, are caused by mutations in essential cellular proteins leading to aggregation or misfolding of the proteins[51–53]. Understanding the mechanisms of these diseases and their subsequent treatment require an understanding of how the relevant proteins fold or misfold. (v) The progress of many genome projects has led to the availability of amino-acid sequences of thousands of proteins. To understand what the roles of many of these proteins in the cell are, it is important to obtain structural information on these proteins. Experimental methods of structure determination (X-ray crystallography and NMR) are slow, and are unable to keep up with the pace of discovery of new proteins. It has, therefore, become imperative to

develop methods, far more reliable than those currently available, for the prediction of structure from amino-acid sequence. The reliability of the prediction methods will improve substantially only after an intimate understanding of the physico-chemical interactions that dictate protein folding is obtained. (vi) Ligand binding, protein–protein interactions and other aspects of protein structure–function relationships will be understood at the molecular level only after the principles governing protein folding are first elucidated.

## Conceptual models

### *The necessity for folding pathways*

The unfolded polypeptide chain of a protein can adopt very many possible conformations. A polypeptide chain with $n+1$ residues, each with an average $m$ equally probable conformations, can have $m^n$ possible conformations. Thermodynamic analysis of protein unfolding[54] suggests that $m \approx 8$ in a real, unfolded polypeptide chain. For a small protein of 100 amino acid residues, this suggests the occurrence of $8^{99}$ or $10^{89}$ conformations. Even if each conformation were sampled for only $10^{-13}$ seconds, which corresponds to the fastest rate of interconversion, it would require more than $10^{66}$ years for the polypeptide chain to sample $10^{89}$ conformations[55]. Real proteins fold, however, within seconds or minutes. This discrepancy between such a calculation and experimental results is known as the 'Levinthal Paradox', and led Levinthal to propose that there must be pathways for protein folding, which speed up the process markedly[56,57]. More recently, it has, however, been shown that if instead of a completely unbiased random search, a small and rational energy bias against locally unfavourable conformations is included in the calculation of the time taken to fold, Levinthal's time is reduced to a biologically meaningful time[55].

### *Framework model*

According to the framework model[58], protein folding is hierarchical: during folding, secondary structural elements are formed initially, which then constitute the framework for subsequent formation of tertiary structure. A more extended interpretation of this model is that when folding commences, fluctuating segments of secondary structure first form and then coalesce so that the polypeptide

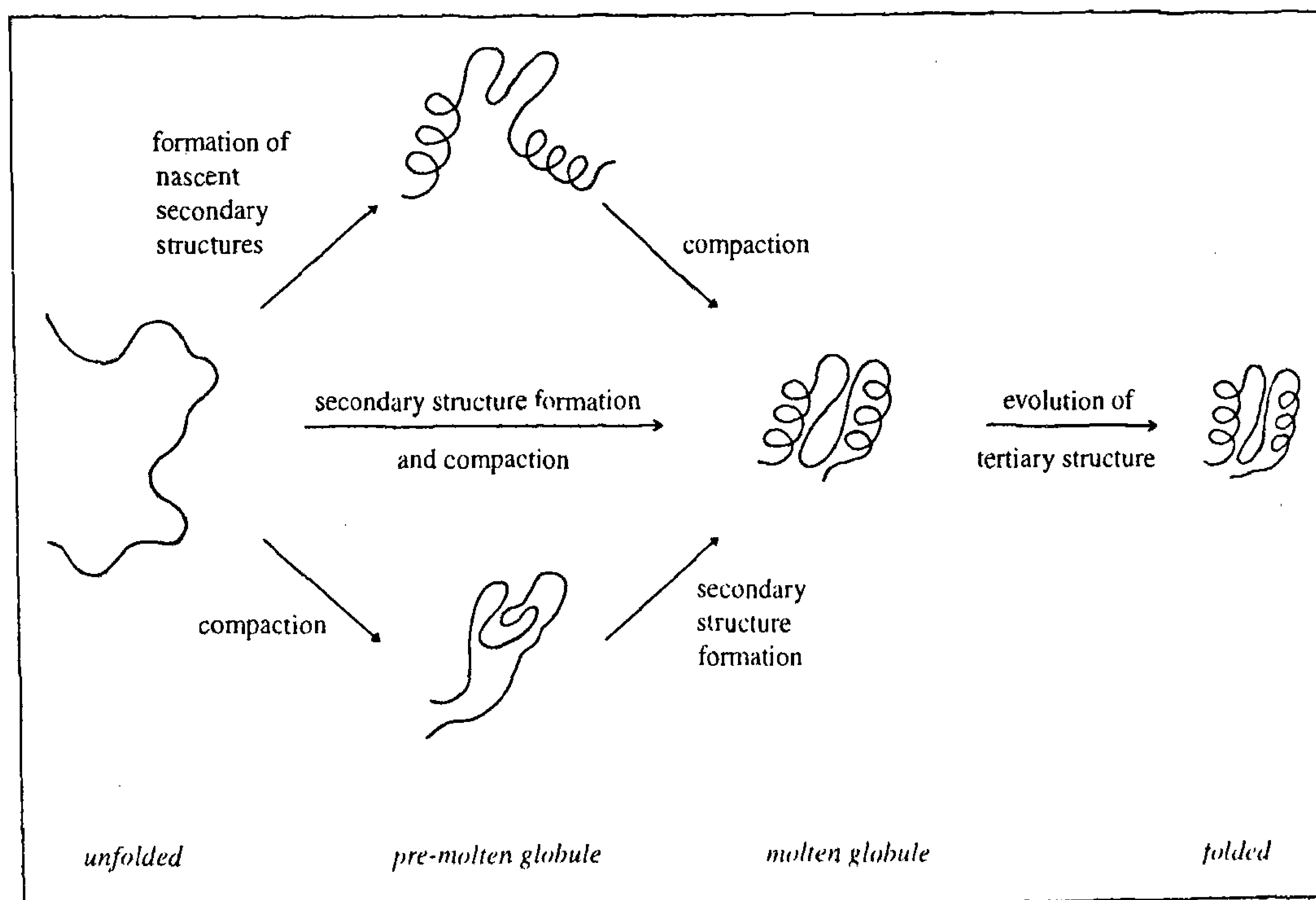


**Figure 1.** Models of protein folding. In the framework model (top pathway) nascent, fluctuating secondary structural elements form initially along the polypeptide chain. These coalesce to form a molten globule intermediate. In this model, the pre-molten globule is less compact, and any secondary structure present is very much more unstable than that in the molten globule. In the hydrophobic collapse model (bottom pathway), a non-specific collapse of the polypeptide chain occurs initially, and this facilitates subsequent formation of secondary structural elements, leading to the formation of the molten globule intermediate. In this model, the pre-molten globule is a disordered (structureless) form that is nearly as compact as the molten globule. The middle pathway depicts both collapse and secondary structure formation occurring simultaneously to yield the molten globule intermediate. In each case, the final tertiary structure evolves slowly from the molten globule intermediate.

chain becomes compact; the final tertiary structure forms subsequently. In part, this model reflects the realization that many proteins appear to have been assembled from segments of secondary structure[59,60]. It would appear that a consequence of the framework model would be that local conformation should depend on the local sequence. This appears, however, not to be true in native globular proteins: the same sequence can very often form different structures[61]. Nevertheless, experimental evidence for the framework model is strong[62–64]. It should be noted that elements of secondary structure ($\alpha$ helices and $\beta$-sheets) may not be stable sub-structures. This is, however, not an essential requirement for their assembly as long as once assembled, the complex of two or more secondary structural elements is more stable than the individual components. The diffusion–collision–adhesion model[65,66] proposed for the assembly of secondary structural units only demands that they be marginally stable and that the complex be more stable.

### *Hydrophobic collapse model*

In the hydrophobic collapse model, the hydrophobic interaction, which can be described intuitively as water driving two non-polar molecules together, so that their contact with water is reduced, is thought to be the dominant interaction in determining the folding of a protein[67,68]. The idea is supported by the observation that the interiors of globular proteins comprise generally of hydrophobic amino acid residues. The non-specific nature of hydrophobic interactions suggested that they could form more quickly than other more specific interactions. This led to the idea of a hydrophobic collapse being the initial step in folding, in which non-polar side-chains are buried in a loose compact structure. Suggestive evidence for a rapid hydrophobic collapse has been found in staphylococcal nuclease, where C-terminal deletion mutant proteins are unfolded but compact and contain substantial secondary structure[69], and in kinetic folding experiments with dihydrofolate reductase (DHFR)[70]. The first direct evidence for the hydrophobic collapse model has been found in the case of barstar, where it has been shown that a non-specific hydrophobic collapse precedes the formation of any secondary structure or native tertiary interactions[71].

### *Nucleation–condensation–growth model*

According to the nucleation–condensation–growth model[57,67,72–74], a region of the polypeptide chain serves as a nucleus for chain propagation to obtain the native state. The nucleus is expected to be unstable by itself, and is therefore present only fleetingly in the denatured state. Since the initial step is rate limiting, no intermediates beyond the nucleus are expected to accumulate in sufficiently detectable quantity, and the folding reaction is approximated by a two-state transition. The nucleation unit must be small enough for random search, and is expected to fully develop only in the transition state of folding. Direct evidence for the nucleation–condensation model is strongest for the folding of chymotrypsin inhibitor 2 (CI2) (ref. 74).

### *Jig-saw puzzle model*

According to this model[75] of protein folding, the assembly of structural elements can occur in many different ways, analogous to the assembly of the pieces of a jig-saw puzzle which can be put together in various ways. There is no definite order to their assembly. In other words, there may not be any unique pathway for the folding of any protein. Recent studies suggest, however, that there are not a large number of equally accessible pathways by which a protein can fold; instead, folding occurs along one or a few well-defined sequential pathways with unique intermediates.

## Development of structure during folding

The most useful way to analyse the pathway by which a protein folds is to characterize structurally and energetically all the intermediates that define the pathway[24]. Although new methodologies have been developed[62–64,76], partly folded intermediates that can be detected in kinetic studies are difficult to study because they have only a transient presence on the kinetic folding pathway. The cooperative nature of folding does not usually allow any intermediate to accumulate to any significant extent, and it is therefore difficult to characterize structurally a kinetic unfolding intermediate. Nevertheless, much information on how structure develops during folding has been obtained recently.

### *Early events in protein folding*

The earliest steps in protein folding are usually studied using rapid kinetic measurements in the millisecond time domain. For many proteins, the earliest events of structure formation take place within the first 2–4 milliseconds, before they can be observed experimentally. The 'burst phase' change in the experimental observable used to monitor folding reports on the very early formation of structure. Information on what structure forms first can also be obtained by examining the structural features of the native protein which might fold faster than the global folding event. For example, a fragment of a protein which can assume native-like structure in isolation

might form first during folding. Residual interactions and structures in the unfolded state may also play important roles in initial structure formation during folding[13,74,77–82]. Predictions have also been made on the early events in protein folding based on known native structures[83–86].

*Formation of secondary structure.* Two powerful probes used to study the formation of early secondary structure are far-UV circular dichroism (CD)-monitored stopped-flow kinetics[87,88] and pulsed hydrogen exchange in conjunction with NMR[19]. Folding of the peptide backbone can be monitored by measuring CD in the far-UV region; for example, formation of $\alpha$-helices can be monitored at 222 nm. Initial formation of secondary structure has been observed for many proteins in far-UV CD-monitored kinetic experiments with millisecond time resolution[88–96]. In most of the stopped-flow CD studies reported so far, a significant amount of signal indicative of secondary structure was observed to have developed within the dead time of the measurement, much faster than the formation of the native structure[90–94]. In some cases, no further changes in ellipticity were found, suggesting that all the native secondary structure is formed in the earliest step[94,95]. In the case of tryptophan synthase, an overshoot of far-UV CD signal was, however, observed which was attributed to transient $\alpha$-helical structures, which may not be native-like[94]. A similar overshoot was also observed in the case of hen lysozyme[90], which was attributed not to the formation of secondary structure, but to transient chain interactions, possibly associated with a disulphide bridge[91]. In some cases, several kinetic phases after the burst phase have been resolved[89,92], which are difficult to assign to the formation of specific structural features[89]. In the case of DHFR, several late phases have been resolved and assigned to the continued development of helices and rearrangement of the central $\beta$-sheet[92].

Whereas CD provides a useful index of secondary structure formation without defining the origin of specific features, pulsed hydrogen exchange labelling methods can detect the formation of specific secondary structure during folding[19,86]. The pulsed hydrogen exchange method, in which the ability of secondary structure to retard the exchange of peptide bond protons to exchange with solvent protons is measured, also allows the determination of the stabilities of the secondary structures formed at various times during the folding process. Early formation of native $\alpha$-helix and $\beta$-sheet has been observed in cytochrome c (ref. 63), RNase A (ref. 62, 97), barnase[64], $\alpha$-lactalbumin[98] and several other proteins[86]. A very early collapsed intermediate on the refolding pathway of the very fast folding form of unfolded RNase A has been shown to possess some marginally stable native-like secondary structural units[99].

Pulsed hydrogen exchange experiments may not, however, detect the initial formation of isolated elements of secondary structure. This has been shown by a comparison of hydrogen exchange data with that from far-UV CD in the cases of hen lysozyme[90] and cytochrome c (ref. 89). The structure formed in the initial stages of folding appeared to be too labile to afford significant protection from the relatively intense labelling pulse used in these experiments[31,99]. Formation of mutually stabilizing clusters of secondary structural elements is probably required to provide protection from exchange. This has been substantiated by studies of peptide fragments of regions of proteins which are protected from exchange initially in folding[100–102]. Thus, the combination of these techniques suggests the formation of abundant, but rapidly fluctuating, secondary structure in the earliest observable folding intermediates of hen lysozyme and cytochrome c.

*Collapse and clustering of side chains.* One of the major questions about very early events in protein folding is whether the initial formation of secondary structure precedes a global hydrophobic collapse of the polypeptide chain or whether the secondary structure is formed as a consequence of such a collapse[103]. This is a difficult question to answer because methods for the determination of a collapse of a polypeptide chain by direct measurements, for example, of the radius of gyration[104], are not readily coupled to rapid data acquisition systems.

Fluorescence is very useful in this context, because of its sensitivity to solvation, and both intrinsic and extrinsic probes have been used to investigate folding reactions. The fluorescence intensity of tryptophan residues is used typically to monitor folding in most kinetic studies because it is sensitive to environment, and its variation can be easily monitored[105,106]. In the case of a protein with a single tryptophan residue, the change in fluorescence can be interpreted in terms of the change in its particular environment[107,108]. Tryptophans have been engineered into various sites of lactate dehydrogenase to study the kinetics of folding of different parts of the structure[109]. A single tryptophan has been engineered into the interior of ubiquitin to monitor hydrophobic clustering[110,111]. In each of these cases, burst phase fluorescence changes within the dead time (5 milliseconds) of the measurement were observed, indicating that at least some side chain interactions involving the novel tryptophan residue had already formed on this time scale. The burst phase changes in intrinsic fluorescence intensity indicate that some degree of solvent exclusion, characteristic of the formation of at least a rudimentary hydrophobic core occurs within 5 milliseconds after commencement of folding.

In the folding pathway of DHFR, two tryptophan residues form rapidly a specific, native-like contact

following an initial burst in secondary structure formation[92]. Fluorescence anisotropy decay measurements used to monitor the binding of a hydrophobic dye, 1-anilino 8-naphthalene sulphonic acid (ANS), during the folding of DHFR showed fluorescence anisotropy decay curves after 20 ms of refolding equivalent to those measured for the native protein[112], indicating that the molecule is similar in size to the native protein at this time point. In the case of hen lysozyme, which has six tryptophans, the fluorescence change within the first few milliseconds of refolding indicates the formation of hydrophobic clustering of at least a few tryptophan residues, although no significant hydrogen exchange protection is found within this time scale[113]. The earliest events in the refolding of lysozyme include both formation of substantial secondary structure, as detected by far-UV CD, and hydrophobic clustering, as indicated by change in fluorescence intensity[114].

Extrinsic probes, like ANS, serve as alternative probes to measure the hydrophobic collapse of a protein during folding[115]. The fluorescence of ANS is quenched in water and increases when it binds to clusters of hydrophobic residues in proteins. For many proteins, ANS fluorescence passes through a maximum during folding when partially folded intermediates accumulate transiently. For several proteins, including $\alpha$-lactalbumin and $\beta$-lactamase, very little enhancement of ANS fluorescence during the first few milliseconds of folding was observed, suggesting that there was no extensive hydrophobic collapse in the initial intermediates[116]. In those experiments ANS was, however, present throughout the folding process, and it has been shown that the binding of ANS to intermediates affects the folding kinetics[117]. A new pulse labelling protocol, in which ANS was added after different times of refolding, has indeed shown that both $\alpha$-lactalbumin and carbonic anhydrase undergo hydrophobic collapse very early during folding[117]. In the case of lysozyme, maximum ANS binding occurs in a burst phase within the dead time, when a substantial tryptophan intensity signal also develops. These data suggest that, in lysozyme, considerable hydrophobic collapse takes place during the initial step of folding, along with secondary structure formation[114]. Time-resolved fluorescence energy transfer measurements have been used to show that the initial folding intermediates of an immunoglobulin light chain[118], barstar[71] and apomyoglobin[119] are compact.

Direct measurements of protein compactness by small angle X-ray scattering have been performed recently both in equilibrium[120] and kinetic[121] experiments. By coupling this technique with rapid mixing, the compactness of an early intermediate in apomyoglobin could be determined in real time, within 100 milliseconds of commencement of folding[121]. Measurement of the scattering intensity spectrum indicates that the protein is indistinguishable in terms of its compactness from the native state much before folding is complete.

Side chain packing at the atomic level during folding has been studied recently by both one-dimensional[122,123] and two-dimensional[124] NMR in real time for systems that fold relatively slowly. Using this technique, the burst-phase intermediate of $\alpha$-lactalbumin has been shown to resemble the acid-induced equilibrium intermediate of the same protein, which is a compact 'molten globule'[123].

The formation of molten globule intermediates is emerging as a common feature on the folding pathways of many proteins. Molten globules possess partial secondary structure stabilized by minimal native-like tertiary interactions, if any. The clustering of hydrophobic residues in these intermediates is monitored easily by their ability to bind hydrophobic dyes such as ANS[115,117,125]. Molten globules were first extensively characterized as equilibrium intermediates[38,116,126–128]. Recent studies have, however, shown that they do actually accumulate on the kinetic folding pathways, very often in a burst phase of a few milliseconds after commencement of folding[98,129]. While it is known that the equilibrium molten globules are compact, the degree of compactness in kinetic molten globules can only be inferred from their equilibrium counterparts.

Spectroscopic techniques that are available currently can measure the structural changes during folding only in the millisecond time scale, because the mixing deadtimes of the mechanical mixing devices are, with very few exceptions[130], typically 1–4 milliseconds. Thus, folding events occurring earlier than that are not observable. Methods are, however, being developed to probe structural events occurring on faster time scales by changing the physical[81,131–134] or chemical[135,136] conditions of the sample to initiate the folding reaction. Temperature jump experiments have been carried out in the case of cold-denatured barstar[81] and apomyoglobin[132] to study folding events in the microsecond timescale. Apomyoglobin forms a compact state within 20 microseconds, involving at least the A- and the H-helix backbones[132,133]. Submillisecond refolding events in the case of cytochrome c have been studied by initiating the refolding reaction either by laser-induced dissociation of a ligand[135] or by rapid photochemical electron transfer[136]. Analysis of the broadened NMR resonances, arising from the fast interconversion of folded and the unfolded species, has been applied to investigate submillisecond folding events in the case of monomeric repressor[137,138].

### *Kinetic barriers and fast folding reactions*

The list of proteins that fold very rapidly ($> 10\,s^{-1}$ at the equilibrium unfolding transition midpoint at 25°C) is growing and includes cold shock protein B (CspB)

(ref. 139), fragments of Streptococcal G protein[140,141] and cytochrome c (ref. 142). The rate of folding of a protein appears not to depend specifically on its size or structure, although in general, larger proteins fold more slowly. There also does not appear to be any correlation between stability and folding rate. Several marginally stable proteins, including CspB and a fragment of Streptococcal G protein[141], fold very rapidly. A mutation that decreases the stability of barstar also increases its rate of folding[143]. RNase A folds slower in conditions where it is more stable, and faster in conditions where it is less stable[144]. On the other hand, a relationship between the rate constant of folding and the stability of the native conformation has been demonstrated in the case of the Arc repressor[145] and ubiquitin[111].

The reasons why folding rates of proteins of similar size vary over a 100-fold range are not understood. Tertiary interactions of native[146] or non-native[142,147] origin sometimes act as kinetic barriers in folding. In the case of cytochrome c, a non-native interaction between the unfolded protein and the heme group is responsible for the slow folding of the protein; elimination of the interaction causes folding to occur 20-fold faster[142]. In the refolding of dimeric Arc repressor, formation of a set of native, partially buried, salt-bridge and hydrogen-bond interactions acts as a kinetic barrier. Replacement of these specific interactions with non-specific hydrophobic interactions accelerates the refolding rate of the protein by three orders of magnitude[146]. Equilibrium molten globules have a tendency to form soluble aggregates in solution, a property likely to be shared by kinetic molten globule intermediates. Disaggregation of a transiently formed kinetic molten globule aggregate on the folding pathway might also act as a rate limiting step during folding.

### *The transition state in protein folding*

The transition state in a protein folding reaction is the structure that is at the highest energy level on the pathway of folding. The classical transition state theory of chemical reactions, proposed by Eyring[148], is applicable to chemical reactions involving the making or breaking of covalent bonds. In contrast, the protein folding reaction involves formation or destruction of mainly weak interactions. Nevertheless, transition state theory appears to be applicable to folding reactions as well. The results of kinetic studies for a few proteins have indicated that the transition state is compact[149] and more native-like than unfolded[150].

Structural information on the transition state of folding can also be extrapolated from characterization of the intermediates that precede the transition state on the folding or the unfolding pathway of a protein. Unfolding and refolding are linked by the principle of microscopic reversibility, requiring that under the same conditions, transition states and intermediates for unfolding and refolding should be identical[151]. In the case of the folding of RNase A in the absence of complications due to non-native interactions, it has, however, been proposed that the transition states on the folding and unfolding pathways are not identical[144].

The advantage of analysing the unfolding pathway is that it starts from the best defined structure on the folding pathway, that of the folded state. In addition, there are no complications arising from slow phases due to proline isomerization. The unfolding kinetics of most proteins have been described by simple two-state transitions, without the involvement of any unfolding intermediates. Recent studies on RNase A by NMR[152], and on barstar by optical probes[153] have, however, demonstrated the accumulation of kinetic intermediates on the unfolding pathway. The kinetic unfolding intermediate of barstar is compact with a dehydrated hydrophobic core and devoid of substantial secondary structure[153]. The unfolding of a low-pH equilibrium intermediate of a mutant form of apomyoglobin shows biphasic kinetics with a large burst phase change in fluorescence[154]. This has, however, been explained in terms of the intermediate being an ensemble of structurally related forms.

The transition state for the folding reaction, in which many hundreds of non-covalent interactions form, may not be unique, but may represent an ensemble of states with very similar energies, or one of several alternative states that are available. Theoretical studies suggest the latter possibility[155], but the demonstration that several proteins show apparent two-state folding/unfolding kinetics[139,156,157] suggests that if multiple transition states are present, they are equivalent in energy.

### *Late events in protein folding*

The folding kinetics of most proteins, with few exceptions[139,157,158], show at least two distinct kinetic phases, usually a fast phase and at least one slow phase, with their time constants significantly different, even when equilibrium studies indicate a simple two-state unfolding transition. As discussed above, major structural changes towards the folded state like the hydrophobic collapse, H-bond formation, the packing of secondary structural elements, etc., appear to take place within the time scale of the fast phase. For many proteins, the reason for the slow processes is the isomerization about proline imide bonds[159-161]. It has been assumed generally that little structure formation accompanies this isomerization process, and for this reason, slow folding reactions have been used as model systems for the quantitative analysis of the coupling of intrinsically slower chemical processes (e.g. imide isomerization) to faster structure-forming

folding reactions[162–164]. The extent of structure formation that occurs in a slow folding reaction has, however, not been well studied for any protein. In particular, the consolidation of tertiary interactions that must occur in this step is poorly understood.

The *X-Pro* imide bond, where *X* is any amino acid residue, is found to be in a *cis* conformation with reasonable frequency (approximately 7% of all prolyl peptide bonds) in many native proteins, whereas *cis* amide bonds have been observed only rarely[165]. In unfolded proteins and in peptides, the *trans* form of the *X-Pro* bond is favoured thermodynamically over its *cis* isomer, and the equilibrium distribution of the two forms are 70–90% and 10–30% for *trans* and *cis*, respectively[166,167]. The presence of these two forms in unfolded proteins gives rise to heterogeneity of the unfolded form[82]. The rates of *cis* ⇔ *trans* interconversions are slow and in the same time range as slow refolding reactions. This led Brandts and his colleagues[168] to propose that the slow phase of protein folding is because of the slow isomerization step from the non-native *trans* form of proline to its native *cis* form during folding. Many model proteins used for folding studies, such as RNase A, RNase T1, thioredoxin and barstar, have *cis* imide linkages, whereas others, such as hen lysozyme, T4 lysozyme, cytochrome c, etc., have only *trans* prolines. The refolding of proteins with one or more native *cis* prolyl imide bonds is usually dominated by slow kinetic phases, while the refolding of proteins containing all *trans* prolyl imide bonds is dominated by fast phases.

Proline isomerization has been implicated directly for being responsible for the slow refolding kinetics of RNase A (refs. 160,169,170), RNase T1 (ref. 159), staphylococcus nuclease[161] and thioredoxin[171]; replacement of *cis* prolines leads to the elimination of the slow-refolding reactions in these proteins. This effect is, however, not seen in the case of human lysozyme[172], indicating that factors other than *cis-trans* isomerization of *X-Pro* bond could also be responsible for slow-folding reactions in some proteins. A cystine knot loop threading mechanism has been suggested to be responsible for the slow phase in the unfolding of human beta-NGF[173].

Two slow phases are found in the refolding of two-disulphide derivatives of bovine pancreatic trypsin inhibitor (BPTI). Using energy calculations, Levitt[174] predicted that the native state was destabilized most strongly in molecules with *Pro8* in the non-native *cis* conformation (all prolyl peptidyl bonds are *trans* in native BPTI). This prediction was later confirmed experimentally[175]; replacing *Pro8* by glutamine in a two-disulphide variant of BPTI abolished the slowest phase of folding. This result implies that the incorrect conformation of a proline residue that destabilizes the native conformation the most, also leads to the strongest deceleration of folding.

## Structural characterization of folding pathways by protein engineering methods

The advent of recombinant DNA techniques, which led to the birth of protein engineering, provided a powerful tool to study the contributions of individual amino acid residues to the conformation and folding of a protein. One of the principal goals of protein engineering methods has been to estimate the informational content of individual amino acid residue positions. The challenge is to determine the contributions of individual residues to the stability and folding of a specific protein. By careful equilibrium and kinetic analysis of mutant proteins, invaluable information can be obtained on the folding pathway of a protein, which is otherwise not available from the studies on the wild type protein.

### *Protein engineering and equilibrium intermediates*

Folding intermediates are populated usually to a negligible extent at equilibrium because they are less stable than the fully folded protein. Thus, many of the small monomeric, single-domain proteins unfold through a highly cooperative process with no detectable accumulation of equilibrium intermediates between the fully folded and the unfolded forms[32]. Site-directed mutagenesis is a potential method for breaking down the cooperativity of the folding process so that folding intermediates accumulate at equilibrium and are more amenable to study[176]. Equilibrium folding intermediates have been populated to a significant extent by perturbing hydrophobic interactions in barnase[177] and by perturbing a buried hydrogen bond in barstar[178] and their structures have been studied spectroscopically.

### *Double mutant cycles*

The extent of interaction between two residues in a folded protein can be estimated by using double mutant cycles[179,180]. The method consists of identifying two interacting residues in the folded protein and mutating them both independently and simultaneously. Then the differences in free energy of unfolding ($\Delta\Delta$Gs) of the three mutant proteins are measured with respect to the wild type protein. The comparison of the difference of the $\Delta\Delta$Gs of the two single mutant proteins with that of the $\Delta\Delta G$ of the double mutant protein indicates the extent of interaction between the two residues under study. The double mutant cycle allows the measurement of the changes in cooperativity along the folding pathway. This approach has been applied to barnase and the results indicated that during the refolding process cooperativity is still present in some regions of the folding intermediate[181].

*Characterizing the transition state by protein engineering methods*

Protein engineering methods offer a glimpse of the structure of the transition state of folding. Structural characterization of the transition state is achieved from the inferences of kinetic measurements on mutant proteins[24,64,151,182–184]. Sufficient experimental progress has now been made to validate this method and the theoretical basis is reasonably well established[185].

The method comprises of: (i) establishing and measuring the interactions that stabilize different parts of the molecule from experiments on the equilibrium stability of mutant proteins and (ii) using kinetic measurements of protein folding and unfolding to determine what fraction of the stabilization energy measured in (i) is used to stabilize the transition state for folding[182]. Each mutation engineered into the protein acts as a reporter group for the progress of folding at the appropriate position. The procedure gives a quantitative parameter, $\Phi$, for the localized comparison of the transition state with the native state[182]. $\Phi$ is calculated as the change in the free energy difference between the transition state and the unfolded state relative to the change in the overall protein stability as a result of the mutation. In situations where a mutation has an identical energetic effect on the folded and transition states, $\Phi = 1$ and the interaction is assumed to be present in the transition state. In cases where $\Phi = 0$, the structure of the transition state in the region is assumed to be identical to that of the unfolded state. Fractional values of $\Phi$, found frequently in the analysis of measurements of the CI2 folding, signify that the structure is weakened or that a dynamic equilibrium exists between folded and unfolded conformations. These fractional values of $\Phi$ are particularly difficult to interpret in structural terms[73,186].

Using the above method, the structures of transition states of several proteins have been characterized. The transition state of CI2 appears to be a globally collapsed state strongly resembling the folded protein but expanded by 30%, and lacking the specific hydrophobic packing characteristic of the native state[186–188]. Structural characterization of the transition state of barnase by protein engineering reveals that it is very similar to the folded state[23,182,184]. The major differences are a weakened main hydrophobic core, the disruption of a small subdomain and of some loops, and the unwinding of the amino termini of the two helices, and less secondary structure. As in the case of CI2 and barnase, considerable secondary structure is formed only after the transition state in folding of Arc repressor[189], CheY (ref. 190) and a fragment from GCN4 (ref. 191). Molecular dynamics simulation studies have also been used to study the structures of the transition states of CI2 (refs 192,193) and barnase[194]. A good correlation between the experimental and the theoretical results[195] corroborates the validity of the procedure of studying transition states by protein engineering.

*Characterizing kinetic folding intermediates by protein engineering*

Although spectroscopic probes such as fluorescence and circular dichroism provide useful information about structure, both at equilibrium and in kinetic studies, they monitor the structure at a macroscopic level. The structural information at the level of individual residues and atoms at equilibrium can be obtained only by NMR. The study of kinetics of structural changes at the atomic level during folding by stopped-flow NMR spectroscopy, which could be the best method to monitor protein folding, is, however, still at an early stage[123–124,152,196].

There are two existing procedures for analysing the structure of the transient intermediates with high structural resolution: pulsed hydrogen exchange[19,86,197] and protein engineering[198]. Although the protein engineering method reports only indirectly on structure, it is more powerful in being applicable to many more regions of the protein, including all the side chains and the hydrophobic core, and it is not just restricted to those NH groups that are protected sufficiently well to be resistant to exchange. Further, the protein engineering method gives information on energetics, and may also be extended to events that occur on shorter time scales than pulsed hydrogen exchange methods[73,81].

Protein engineering methods have been used to characterize transient intermediates on the folding pathways of barnase[198,199] and phosphoglycerate kinase[200]. The transient folding intermediate of barnase is characterized by relative positions of $\alpha$-helices and $\beta$-sheet being native like, as well as the relative positions of $\beta$-strands within the sheet being native-like. The structure looks very much like the transition state, except that the intermediate has a weaker hydrophobic core[198,199]. The structure of the transient intermediate determined from protein engineering studies agrees well with that obtained from hydrogen exchange studies.

*Engineering disulphide bonds to study protein folding*

Protein folding has been studied by kinetic analysis of disulphide bond formation in proteins like BPTI, which requires formation of specific disulphide bonds for its folding[201]. Folding pathways of such proteins have been studied and trapped intermediates have been characterized by selectively mutating cysteine residues and measuring the rate of disulphide bond formation[202]. Engineered disulphide bonds have been used to report on the folding

of specific regions of a protein[203], where formation of a disulphide bond is not required for folding. The engineered disulphide bond is expected to stabilize an early kinetic intermediate if the bond is located in the region which folds early in the pathway.

The introduction of free cysteine residues in phosphoglycerate kinase and the analysis of their reactivities during the folding reaction has also provided useful structural information about kinetic intermediates[204]. These studies suggest that the kinetic folding intermediates contain an ordered native-like secondary structure, especially near the centre of the protein, that is weakened at the periphery.

## Concluding remarks

Experimental studies of protein folding are being provided with valuable insight from simplified statistical models[36,37,42]. Theoretical studies have now become sufficiently sophisticated to be able to describe many of the basic features of folding that are evident from experimental studies[155,192–195,205–208]. Clearly, the study of protein folding will benefit tremendously from the marriage of theoretical and experimental studies.

The results of experimental studies with millisecond time resolution on the folding pathways of many proteins, including RNase A (ref. 62,99), $\alpha$-lactalbumin[98], lysozyme[30], apomyoglobin[129], cytochrome c (refs 89,135) and DHFR[95,112] suggest that compaction of the polypeptide chain and secondary structure formation occur simultaneously, because these two processes cannot be resolved temporally. It remains to be seen, whether the two processes indeed occur simultaneously in a faster time domain. Folding experiments have only recently entered the microsecond and nanosecond time domain, and the possibility of nanosecond circular dichroism measurements[209] holds great promise. Future studies will provide crucial information on how folding commences. In the case of barstar, it has been shown that a hydrophobic collapse precedes secondary structure formation[71], leading to the formation of a disordered (structure less) globule, and recent temperature jump experiments in the microsecond time scale[81] appear to support this result.

Partially folded forms of a protein can often be detected in equilibrium with the fully folded form, and much recent effort has gone into obtaining structural information on these forms[39], and relating them to structural intermediates that accumulate transiently on the direct folding pathway. In particular, hydrogen exchange measurements of proteins in the presence of mildly destabilizing concentrations of denaturants[210–214] yield structural information on partly unfolded forms of the protein in rapid equilibrium with the fully folded protein. These partly folded forms may or may not, however, represent structural intermediates on the direct folding pathway[213]. While in the case of cytochrome c, the partly unfolded forms do resemble structural intermediates that have been identified on the kinetic folding pathway[211], in the case of barnase, the partly unfolded form does not resemble the kinetic intermediate that has been structurally characterized[213]. It is essential to determine if partly folded forms of a protein, in equilibrium with either the unfolded state (in kinetic experiments) or with the folded state (in hydrogen exchange experiments), are tangibly present on the direct folding pathway. It has been shown recently that a non-native intermediate accumulates during the folding of $\beta$-lactoglobulin[215]. This has again highlighted the importance of demonstrating whether partly folded intermediates, which are observed in kinetic experiments, accumulate on the direct folding pathway or whether they are off-pathway dead-end intermediates[20].

Although it has now been clearly established that many proteins fold along one or more well-defined pathways, which involve intermediate structures that are only partly folded[16,27], recent studies have shown that the folding/unfolding of small proteins, such as CspB (ref. 139), CI2 (ref. 156), the IgG binding domain of streptococcal protein G (ref. 141), acyl-coenzyme binding protein[156], the SH3 domain of spectrin[217] and cytochrome c (refs. 142, 216) follow two-state kinetics: no intermediates are detectable between the unfolded and the folded state[218]. These observations have raised the question whether intermediates are essential because they guide the folding process along a pathway and thereby speed up the folding process, or whether they are only incidental to the folding process and accumulate simply because of the presence of kinetic barriers. Do intermediates that are too unstable to be optically detectable exist on the folding pathways of proteins that fold by apparent two-state behavior, where they function in a spectroscopically silent manner? Does the accumulation of a stable intermediate on a folding pathway actually cause folding to proceed slower than it would in its absence[64,111,120,144,219]? Identification of the natures of kinetic barriers that retard folding, and their structural description are major goals in the study of protein folding.


1. Anson, M. L. and Mirsky, A. E., *J. Gen. Physiol.*, 1925, **9**, 169–179.
2. Anson, M. L. and Mirsky, A. E., *J. Phys. Chem.*, 1939, **35**, 185.
3. Anson, M. L., *Adv. Protein Chem.*, 1945, **2**, 361.
4. Lumry, R. and Eyring, H., *J. Phys. Chem.*, 1954, **58**, 110–120.
5. Anfinsen, C. B., Haber, E., Sela, M. and White, F. H., *Proc. Natl. Acad. Sci. USA*, 1961, **47**, 1309–1314.
6. Epstein, C. J., Goldberger, C. B. and Anfinsen, C. B., *Cold Spring Harbor Symp. Quant. Biol.*, 1963, **28**, 439.
7. Anfinsen, C. B., *Science*, 1973, **181**, 223–230.
8. Jaenicke, R., *Prog. Biophys. Mol. Biol.*, 1987, **49**, 117–237.
9. Creighton, T. E., *Proc. Natl. Acad. Sci. USA*, 1988, **85**, 5082–5086.
10. Alber, T., *Annu. Rev. Biochem.*, 1989, **58**, 765–798.
11. King, J., *Chem. Eng. News*, 1989, **67**, 32–54.

12. Shortle, D., *Adv. Protein Chem.*, 1995, **46**, 217–248.
13. Shortle, D., *Curr. Opin. Struct. Biol.*, 1996, **6**, 24–30.
14. Bowie, J. U., Reidhaar-Olson, J. F., Lim, W. A. and Sauer, R. T., *Science*, 1990, **247**, 1306–1310.
15. Creighton, T. E., *Biochem. J.*, 1990, **270**, 1–16.
16. Kim, P. S. and Baldwin, R. L., *Annu. Rev. Biochem.*, 1990, **59**, 631–660.
17. Dill, K. A. and Shortle, D., *Annu. Rev. Biochem.*, 1991, **60**, 795–825.
18. Kuwajima, K., *Curr. Opin. Biotechnol.*, 1992, **3**, 462–467.
19. Baldwin, R. L., *Curr. Opin. Struct. Biol.*, 1993, **3**, 84–91.
20. Baldwin, R. L., *Fold. Des.*, 1996, **1**, R1–R8.
21. Dyson, H. J. and Wright, P. E., *Annu. Rev. Biophys. Chem.*, 1991, **20**, 519–538.
22. Dyson, H. J. and Wright, P. E., *Curr. Opin. Struct. Biol.*, 1993, **3**, 60–65.
23. Fersht, A. R., *FEBS Lett.*, 1993, **325**, 5–16.
24. Fersht, A. R., *Curr. Opin. Struct. Biol.*, 1994, **5**, 79–84.
25. Rose, G. D. and Wolfenden, R., *Annu. Rev. Biophys. Biomol. Struct.*, 1993, **22**, 381–415.
26. Scholtz, J. M. and Baldwin, R. L., *Annu. Rev. Biochem.*, 1993, **21**, 95–118.
27. Matthews, C. R., *Annu. Rev. Biochem.*, 1993, **62**, 653–683.
28. Matthews, B. W., *Adv. Protein Chem.*, 1995, **46**, 249–278.
29. Richards, F. M. and Lim, W. A., *Q. Rev. Biophys.*, 1994, **26**, 423–498.
30. Dobson, C. M., *Curr. Biol.*, 1994, **4**, 636–640.
31. Evans, P. A. and Radford, S. E., *Curr. Opin. Struct. Biol.*, 1994, **4**, 100–106.
32. Privalov, P. L., in *Protein Folding* (ed. Creighton, T. E.), W. H. Freeman & Co., New York, 1992, pp. 83–126.
33. Ptitsyn, O. B., in *Protein Folding* (ed. Creighton, T. E.), W. H. Freeman & Co., New York, 1992, pp. 243–300.
34. Ptitsyn, O. B., *Protein Eng.*, 1994, **7**, 593–596.
35. Roder, H. and Elöve, G. A., in *Mechanism of Protein Folding* (ed. Pain, R. H.), IRL Press, New York, 1994, pp 26–54.
36. Dill, K., Bromberg, S., Yue, K., Fiebig, K. M., Yee, D. P., Thomas, P. D. and Chan, H. S., *Protein Sci.*, 1995, **4**, 561–602.
37. Karplus, M. and Šali, A., *Curr. Opin. Struct. Biol.*, 1995, **5**, 58–73.
38. Fink, A. L., *Annu. Rev. Biophys. Biomol. Struct.*, 1995, **24**, 495–522.
39. Shortle, D., Wang, Y., Gillespie, J. R. and Wrabl, J. O., *Protein Sci.*, 1996, **5**, 991–1000.
40. Cordes, M. H. J., Davidson, A. R. and Sauer, R. T., *Curr. Opin. Struct. Biol.*, 1996, **6**, 3–10.
41. Miranker, A. D. and Dobson, C. M., *Curr. Opin. Struct. Biol.*, 1996, **6**, 31–42.
42. Shakhnovich, E. I., *Fold. Des.*, 1996, **1**, R50–R54.
43. Terwilliger, T. C., *Adv. Protein Chem.*, 1995, **46**, 177–215.
44. Hartl, F. U., *Nature*, 1996, **381**, 571–580.
45. Thatcher, D. R. and Hitchcock, A., in *Mechanisms of Protein Folding* (ed. Pain, R. H.), IRL Press, New York, 1994, pp. 229–261.
46. Prouty, W. F., Karnovsky, M. J. and Goldberg, A. L., *J. Biol. Chem.*, 1975, **250**, 1112–1122.
47. O'Shea, E. K., Lumb, K. J. and Kim, P. S., *Curr. Biol.*, 1993, **3**, 658–667.
48. Quinn, T. P., Tweedy, N. B., Williams, R. W., Richardson, J. S. and Richardson, D. C., *Proc. Natl. Acad. Sci. USA*, 1994, **91**, 8747–8751.
49. Tanaka, T., Kuroda, Y., Kimura, H., Kikidoro, S.-I. and Nakamura, H., *Protein Eng.*, 1994, **7** 969–976.
50. Raleigh, D. P., Betz, S. F. and DeGrado, W. F., *J. Am. Chem. Soc.*, 1995, **117**, 7558–7559.
51. Pepys, M. B., Hawkins, P. N., Booth, D. R. *et al.*, *Nature*, 1993, **362**, 553–557.
52. Thomas, P. J., Qu, B.-H. and Pederson, P. L., *Trends Biochem. Sci.*, 1995, **20**, 456–459.
53. Blacklow, S. C. and Kim, P. S., *Nature Struct. Biol.*, 1996, **3**, 758–761.
54. Privalov, P. L., *Adv. Protein Chem.*, 1979, **33**, 167.
55. Zwanzig, R., Szabo, A. and Bagchi, B., *Proc. Natl. Acad. Sci. USA*, 1992, **89**, 20–22.
56. Levinthal, C., *J. Chem. Phys.*, 1968, **65**, 44–45.
57. Levinthal, C., in *Mössbauer Spectroscopy in Biological Systems* (eds Debrunner, P., Tsibris, J. C. M. and Münck, E.), University of Illinois Press, Urbana, 1969, pp. 22–24.
58. Ptitsyn, O. B. and Rashin, A. A., *Biophys. Chem.*, 1975, **3**, 1–20.
59. Levitt, M. and Chothia, C., *Nature*, 1976, **261**, 552–558.
60. Richmond, T. J. and Richards, F. M., *J. Mol. Biol.*, 1978, **119**, 537–555.
61. Kabsch, W. and Sander, C., *Proc. Natl. Acad. Sci. USA*, 1984, **81**, 1075–1078.
62. Udgaonkar, J. B. and Baldwin, R. L., *Nature*, 1988, **335**, 694–699.
63. Roder, H., Elöve, G. A. and Englander, S. W., *Nature*, 1988, **335**, 700–704.
64. Matouschek, A., Kellis, J. T. Jr., Serrano, L., Bycroft, M. and Fersht, A. R., *Nature*, 1990, **346**, 440–445.
65. Karplus, M. and Weaver, D. L., *Nature*, 1976, **260**, 404–406.
66. Karplus, M. and Weaver, D. L., *Protein Sci.*, 1994, **3**, 650–668.
67. Matheson, R. R. and Scheraga, H. A., *Macromolecules*, 1978, **11**, 819–829.
68. Dill, A. K., *Biochemistry*, 1990, **29**, 7133–7155.
69. Shortle, D. and Meeker, A. K., *Biochemistry*, 1989, **28**, 936–944.
70. Garvey, E. P., Swank, J. and Matthews, C. R., *Proteins: Struc. Func. Genet.*, 1989, **6**, 259.
71. Agashe, V. R., Shastry, M. C. R. and Udgaonkar, J. B., *Nature*, 1995, **377**, 754–757.
72. Wetlaufer, D. B., *Proc. Natl. Acad. Sci. USA*, 1973, **70**, 697–701.
73. Fersht, A. R., *Proc. Natl. Acad. Sci. USA*, 1995, **92**, 10869–10873.
74. Neira, J. L., Davis, B., Ladurner, A. G., Buckle, A. M., de Prat Gay, G. and Fersht, A. R., *Fold. Des.*, 1996, **1**, 189–208.
75. Harrison, S. C. and Durbin, R., *Proc. Natl. Acad. Sci. USA*, 1985, **82**, 4028–4030.
76. Plaxco, K. W. And Dobson, C. M., *Curr. Opin. Struct. Biol.*, 1996, **6**, 630–636.
77. Evans, P. A., Topping, K. D., Woolfson, D. N. and Dobson, C. M., *Proteins*, 1991, **9**, 248–266.
78. Neri, D., Billeter, M., Wider, G. and Wüthrich, K., *Science*, 1992, **257**, 1559–1563.
79. Shortle, D., *Curr. Opin. Struct. Biol.*, 1993, **3**, 66–74.
80. Freund, S. M., Wong, K.-B. and Fersht, A. R., *Proc. Natl. Acad. Sci. USA*, 1996, **93**, 10600–10603.
81. Nölting, B., Golbik, R. and Fersht, A. R., *Proc. Natl. Acad. Sci. USA*, 1995, **92**, 10668–10672.
82. Houry, W. A. and Scheraga, H. A., *Biochemistry*, 1996, **35**, 11719–11733.
83. Chelvanayagam, G., Reich, Z., Bringas, R. and Argos, P., *J. Mol. Biol.*, 1992, **227**, 901–916.
84. Rooman, M. J. and Wodan, S. J., *Biochemistry*, 1992, **31**, 10239–10249.
85. Kim, K.-S., Fuchs, J. A. and Woodward, C. K., *Biochemistry*, 1993, **32**, 9600–9608.
86. Woodward, C. K., *Curr. Opin. Struct. Biol.*, 1994, **4**, 112–116.
87. Labhardt, A. M., *Methods Enzymol.*, 1986, **131**, 126–135.
88. Kuwajima, K., Yamaya, H., Miwa, S., Sugai, S. and Nagamura, T., *FEBS Lett.*, 1987, **221**, 115–118.
89. Elöve, G. A., Chaffotte, A. F., Roder, H. and Goldberg, M. E., *Biochemistry*, 1992, **31**, 6876–6883.
90. Radford, S. E., Dobson, C. M. and Evans, P. A., *Nature*, 1992, **358**, 302–307.
91. Chaffotte, A., Guillou, Y. and Goldberg, M. E., *Biochemistry*, 1992, **31**, 9694–9702.

92. Kuwajima, K., Garvey, E. P., Finn, B. E., Matthews, C. R. and Sugai, S., *Biochemistry*, 1991, 30, 7693–7703.
93. Sugawara, T., Kuwajima, K. and Sugai, S., *Biochemistry*, 1991, **30**, 2698–2706.
94. Chaffotte, A. F., Cadieux, C., Guillou, Y. and Goldberg, M. E., *Biochemistry*, 1992, **31**, 4303–4308.
95. Mann, C. J. and Matthews, C. R., *Biochemistry*, 1993, **32**, 5282–5290.
96. Sugawara, T., Kuwajima, K. and Sugai, S., *Biochemistry*, 1991, **30**, 2698–2706.
97. Udgaonkar, J. B. and Baldwin, R. L., *Proc. Natl. Acad. Sci. USA*, 1990, **87**, 8197–8201.
98. Arai, M. and Kuwajima, K., *Fold. Des.*, 1996, **1**, 275–287.
99. Houry, W. A. and Scheraga, H. A., *Biochemistry*, 1996, **35**, 11734–11746.
100. Sancho, J., Neira, J. L. and Fersht, A. R., *J. Mol. Biol.*, 1992, **224**, 749–758.
101. Wu, L. C., Laub, P. B., Elöve, G. A., Carey, J. and Roder, H., *Biochemistry*, 1993, **32**, 10271–10276.
102. Waltho, J. P., Feher, V. A., Marutka, G., Dyson, H. J. and Wright, P. E., *Biochemistry*, 1993, **32**, 6337–6347.
103. Dill, K. A., Fiebig, K. M. and Chan, H. S., *Proc. Natl. Acad. Sci. USA*, 1993, **90**, 1942–1946.
104. Lattman, E. E., *Curr. Opin. Struct. Biol.*, 1994, **4**, 87–92.
105. Eftink, M. R., in *Topics in Fluorescence Spectroscopy* (ed. Lakowicz, J. R.), Plenum Press, New York, 1991, vol. 2, pp. 53–120.
106. Eftink, M. R., *Biophys. J.*, 1994, **66**, 482–501.
107. Kiefhaber, T., Schmid, F. X., Willaert, K., Engelborghs, Y. and Chaffotte, A., *Protein. Sci.*, 1992, **1**, 1162–1172.
108. Varley, P., Gronenborn, A. M., Christensen, H., Wingfield, P. T., Pain, R. H. and Clore, G. M., *Science*, 1993, **260**, 1110–1113.
109. Smith, C. J., Clarke, A. R., Chia, W. N., Irons, W. L., Atkinson, T. and Holbrook, J. J., *Biochemistry*, 1991, **30**, 1028–1036.
110. Khorasanizadeh, S., Peters, I. D. , Butt, T. R. and Roder, H., *Biochemistry*, 1993, **32**, 7054–7063.
111. Khorasanizadeh, S., Peters, I. D., and Roder, H., *Nature Struct. Biol.*, 1996, **3**, 193–205.
112. Jones, B. E., Beecham, J. M. and Matthews, C. R., *Biochemistry*, 1995, **34**, 1867–1877.
113. Itzhaki, L. S., Evans, P. A., Dobson, C. M. and Radford, S. E., *Biochemistry*, 1994, **33**, 5212–5220.
114. Dobson, C. M., Evans, P. A. and Radford, S. E., *Trends Biochem. Sci.*, 1994, **19**, 31–37.
115. Semisotnov, G. V., Rodionova, N. A., Razgulyaev, O. I., Uversky, V. N., Gripas, A. F. and Golmanshin, R. I., *Biopolymers*, 1991, **31**, 119–128.
116. Ptitsyn, O. B., Pain, R. H., Semisotnov, G. V., Zerovnik, E. and Razgulyaev, O. I., *FEBS Lett.*, 1990, **262**, 20–24.
117. Engelhard, M. and Evans, P. A., *Protein Sci.*, 1995, **4**, 1553–1562.
118. Kawata, Y. And Hamguchi, K., *Biochemistry*, 1991, **30**, 4367–4373.
119. Rischel, C. and Poulsen, F. M., *FEBS Lett.*, 1995, **374**, 105–109.
120. Kataoka, M., Nishii, I., Fujisawa, T., Ueki, T., Tokunaga, F. and Goto, Y., *J. Mol. Biol.*, 1995, **249**, 215–228.
121. Eliezer, D., Jennings, P. A., Wright, P. E., Doniach, S., Hodgson, K. O. and Tsuruta, H., *Science*, 1995, **270**, 487–488.
122. Frieden, C., Hoeltzli, S. D. and Ropson, I. J., *Protein Sci.*, 1993, **2**, 2007–2014.
123. Balbach, J., Forge, V., Van Nuland, N. A. J., Winder, S. L., Hore, P. J. and Dobson, C. M., *Nature Struct. Biol.*, 1995, **2**, 865–870.
124. Balbach, J., Forge, V., Lau, W. S., van Nuland, N. A. J., Brew, K. and Dobson, C. M., *Science*, 1996, **274**, 1161–1163.
125. Shastry, M. C. R. and Udgaonkar, J. B., *J. Mol. Biol.*, 1995, **247**, 1013–1027.
126. Ptitsyn, O. B., *J. Protein Chem.*, 1987, **6**, 273–293.
127. Kuwajima, K., *Proteins: Struct. Func. Genet.*, 1989, **6**, 87–103.
128. Hughson, F. M., Wright, P. E. and Baldwin, R. L., *Science*, 1990, **249**, 1544–1548.
129. Jennings, P. A. and Wright, P. E., *Science*, 1993, **262**, 892–895.
130. Chan, C.-K., Hu, Y., Takahashi, S., Rousseau, D. L., Eaton, W. A. and Hofrichter, J., *Biophys. J.*, 1996, **70**, A177 (abstr.).
131. Phillips, C. M., Mizutani, Y. and Hochstrasser, R. M., *Proc. Natl. Acad. Sci. USA*, 1995, **92**, 7292–7296.
132. Ballew, R. M., Sabelko, J. and Gruebele, M., *Proc. Natl. Acad. Sci. USA*, 1996, **93**, 5759–5764.
133. Ballew, R. M., Sabelko, J. and Gruebele, M., *Nature Struct. Biol.*, 1996, **3**, 923–926.
134. Williams, S., Causgrove, T. P., Gilmanshin, R., Fang, K. S., Callender, R. H., Woodruff, W. H. and Dyer, R. B., *Biochemistry*, 1996, **35**, 691–697.
135. Jones, C. M., Henry, E. R., Hu, Y. *et al.*, *Proc. Natl. Acad. Sci. USA*, 1993, **90**, 11860–11864.
136. Pascher, T., Chesick, J. P., Winkler, J. R. and Gray, H. B., *Science*, 1996, **271**, 1558–1560.
137. Huang, G. S. and Oas, T. G., *Proc. Natl. Acad. Sci. USA*, 1995, **92**, 6878–6882.
138. Burton, R. E., Huang, G. S., Daugherty, M. A., Fullbright, P. W. and Oas, T. G., *J. Mol. Biol.*, 1996, **263**, 311–322.
139. Schindler, T., Herrler, M., Marahiel, M. A. and Schimd, F. X., *Nature Struct. Biol.*, 1995, **2**, 663–673.
140. Kuszwewski, J., Clore, G. M. and Gronenborn, A. M., *Protein Sci.*, 1994, **3**, 1945–1952.
141. Alexander, P., Orban, J. and Bryan, P., *Biochemistry*, 1992, **31**, 7243–7248.
142. Sosnick, T. R., Mayne, L., Hiller, R. and Englander, S. W., *Nature Struct. Biol.*, 1994, **1**, 149–156.
143. Nath, U. and Udgaonkar, J. B., (communicated), 1997.
144. Houry, W. A., Rothwarf, D. M. And Scheraga, H. A., *Nat. Struct. Biol.*, 1995, **2**, 495–503.
145. Milla, M. E., Brown, B. M., Waldburger, C. D. and Sauer, R. T., *Biochemistry*, 1995, **34**, 13914–13919.
146. Waldburger, C. D., Jonsson, T. and Sauer, R., *Proc. Natl. Acad. Sci. USA*, 1996, **93**, 2629–2634.
147. Kiefhaber, T., Grunert, H. P., Hahn, U. and Schmid, F. X., *Prot. Struct. Func. Genet.*, 1992, **12**, 171–179.
148. Eyring, H., *Chem. Rev.*, 1935, **17**, 65.
149. Segawa, S. and Sugihara, M., *Biopolymers*, 1984, **23**, 2473–2488.
150. Kuwajima, K., Mitani, M. and Sugai, S., *J. Mol. Biol.*, 1989, **206**, 547–561.
151. Serrano, L., *Curr. Opin. Struct. Biol.*, 1994, **4**, 107–111.
152. Kiefhaber, T., Labhardt, A. M. and Baldwin, R. L., *Nature*, 1995, **375**, 513–515.
153. Nath, U., Agashe, V. R. and Udgaonkar, J. B., *Nature Struct. Biol.*, 1996, **3**, 920–923.
154. Jamin, M. and Baldwin, R. L., *Nature Struct. Biol.*, 1996, **3**, 613–618.
155. Šali, A., Shakhnovich, E. and Karplus, M., *Nature*, 1994, **369**, 248–251.
156. Jackson, S. E. and Fersht, A. R., *Biochemistry*, 1991, **30**, 10428–10435.
157. Kragelund, B. B., Robinson, C. V., Knudsen, J. and Dobson, C. M., *Biochemistry*, 1995, **34**, 7217–7224.
158. Jackson, S. E., Elmasry, N. and Fersht, A. R., *Biochemistry*, 1993, **32**, 11270–11278.
159. Kiefhaber, T., Grunert, H. P., Hahn, U. and Schimd, F. X., *Biochemistry*, 1990, **29**, 6475–6480.
160. Schultz, D. A., Schimd, F. X. and Baldwin, R. L., *Protein Sci.*, 1992, **1**, 917–924.
161. Nakano, T., Antonino, L. C., Fox, R. O. and Fink, A. L., *Biochemistry*, 1993, **32**, 2534–2541.

162. Hagerman, P. J. and Baldwin, R. L., *Biochemistry*, 1976, **15**, 1462–1473.
163. Kiefhaber, T., Kohler, H. H. and Schimd, F. X., *J. Mol. Biol.*, 1992, **224**, 217–220.
164. Kiefhaber, T. and Schmid, F. X., *J. Mol. Biol.*, 1992, **224**, 231–240.
165. Stewart, D. E., Sarkar, A. and Wampler, J. E., *J. Mol. Biol.*, 1990, **214**, 253–260.
166. Cheng, H. N. and Bovey, F. A., *Biopolymers*, 1977, **16**, 1465–1472.
167. Grathwahl, C. and Wöthrich, K., *Biopolymers*, 1981, **20**, 2623–2633.
168. Brandts, J. F., Halvorson, H. R. and Brennan, M., *Biochemistry*, 1975, **14**, 4953–4963.
169. Houry, W. A., Rothwarf, D. M. And Scheraga, H. A., *Biochemistry*, 1994, **33**, 2516–2530.
170. Dodge, R. W. and Scheraga, H. A., *Biochemistry*, 1996, **35**, 1548–1559.
171. Kelley, R. F. and Richards, F. M., *Biochemistry*, 1987, **26**, 6765–6775.
172. Herning, T., Yutani, K., Taniyama, Y. and Kikuchi, M., *Biochemistry*, 1991, **30**, 9882–9891.
173. De Young, L. R., Burton, L. E., Liu, J., Powell, M. F., Schmelzer, C. H. and Skelton, N. J., *Protein Sci.*, **5**, 1554–1566.
174. Levitt, M., *J. Mol. Biol.*, 1981, **145**, 251–264.
175. Hurle, M. R., Anderson, S. and Kuntz, I. D., *Protein Eng.*, 1991, **4**, 451–455.
176. Zitzewitz, J. A. and Matthews, C. R., *Curr. Opin. Struct. Biol.*, 1993, **3**, 594–600.
177. Sanz, J. M. and Fersht, A. R., *Biochemistry*, 1993, **32**, 13584–13592.
178. Nath, U. and Udgaonkar, J. B., *Biochemistry*, 1995, **34**, 1702–1713.
179. Carter, P. J., Winter, G., Wilkinson, A. J. and Fersht, A. R., *Cell*, 1984, **38**, 835–840.
180. Serrano, L., Horovitz, A., Avron, B., Bycroft, M. and Fersht, A. R., *Biochemistry*, 1992, **29**, 9343–9352.
181. Horovitz, A. and Fersht, A. R., *J. Mol. Biol.*, 1992, **224**, 733–740.
182. Matouschek, A., Kellis, J. T. Jr., Serrano, L. and Fersht, A. R., *Nature*, 1989, **342**, 122–126.
183. Matouschek, A. and Fersht, A. R., *Methods Enzymol.*, 1991, **202**, 81–112.
184. Serrano, L., Matouschek, A. and Fersht, A. R., *J. Mol. Biol.*, 1992, **224**, 805–818.
185. Fersht, A. R., Matouschek, A. and Serrano, L., *J. Mol. Biol.*, 1992, **224**, 771–782.
186. Fersht, A. R., Itzhaki, L. S., Elmasry, N., Matthews, J. M. and Otzen, D. E., *Proc. Natl. Acad. Sci. USA*, 1994, **91**, 10426–10429.
187. Otzen, D. E., Itzhaki, L. S., Elmasry, N. F., Jachson, S. E. and Fersht, A. R., *Proc. Natl. Acad. Sci. USA*, 1994, **91**, 10422–10425.
188. Itzhaki, L. S., Otzen, D. E. and Fersht, A. R., *J. Mol. Biol.*, 1995, **254**, 260–288.
189. Schildach, J., Milla, M., Jeffrey, P., Raumann, B. and Sauer, R., *Biochemistry*, 1995, **34**, 1405–1412.
190. López-Hernàndez, E. and Serrano, L., *Fold. Des.*, 1996, **1**, 43–55.
191. Sosnick, T. R., Jackson, S., Englander, S. W. and DeGrado, W., *Proteins*, 1996, **24**, 427–432.
192. Li, A. and Daggett, V., *Proc. Natl. Acad. Sci. USA*, 1994, **91**, 10430–10434.
193. Li, A. and Dagget, V., *J. Mol. Biol.*, 1996, **257**, 412–429.
194. Caflisch, A. and Karplus, M., *Proc. Natl. Acad. Sci. USA*, 1994, **91**, 1746–1750.
195. Dagget, V., Li, A., Itzhaki, L. S., Otzen, D. E. and Fersht, A. R., *J. Mol. Biol.*, 1996, **257**, 430–440.
196. Hoeltzli, S. D. and Frieden, C., *Proc. Natl. Acad. Sci. USA*, 1995, **92**, 9318–9322.
197. Englander, S. W. and Mayne, L., *Annu. Rev. Biophys. Biomol. Struct.*, 1992, **21**, 243–265.
198. Matouschek, A., Serrano, L. and Fersht, A. R., *J. Mol. Biol.*, 1992, **224**, 819–835.
199. Serrano, L., Matouschek, A. and Fersht, A. R., *J. Mol. Biol.*, 1992, **224**, 847–859.
200. Parker, M. J., Sessions, R. B., Badcoe, I. G. and Clarke, A. R., *Fold. Des.*, 1996, **1**, 145–156.
201. Creighton, T. E., in *Protein Folding* (ed. Creighton, T. E.), W. H. Freeman & Co., New York, 1992, pp. 301–352.
202. Darby, N. J. and Creighton, T. E., *J. Mol. Biol.*, 1993, **232**, 873–896.
203. Clarke, J. and Fersht, A. R., *Biochemistry*, 1993, **32**, 4322–4329.
204. Ballery, N., Desmadril, M., Minard, P. and Yon, J. M., *Biochemistry*, 1993, **32**, 708–714.
205. Abkevich, V., Gutin, A. and Shakhnovich, E., *Biochemistry*, 1994, **33**, 10026–100036.
206. Gutin, A. M., Abkevich, V. and Shakhnovich, E., *Biochemistry*, 1995, **34**, 3066–3076.
207. Wolynes, P. G., Onuchic, J. N. and Thiramulai, D., *Science*, 1995, **267**, 1619–1620.
208. Shakhnovich, E., Abkevich, V. and Ptitsyn, O., *Nature*, 1996, **379**, 96–98.
209. Zhang, C. -F., Lewis, J. W., Cerpa, R., Kunz, I. D. and Kliger, D. S., *J. Phys. Chem.*, 1993, **97**, 5499–5505.
210. Mayo, S. L. and Baldwin, R. L., *Science*, 1993, **262**, 873–876.
211. Bai, Y., Sosnick, T. R., Mayne, L. and Englander, S. W., *Science*, 1995, **269**, 192–197.
212. Chamberlain, A. K., Handel, T. M. and Marqusee, S., *Nature Struct. Biol.*, 1996, **3**, 782–787.
213. Clarke, J. and Fersht, A. R., *Fold. Des.*, 1996, **1**, 243–254.
214. Bhuyan, A. and Udgaonkar, J. B., (communicated), 1997.
215. Hamada, D., Segawa, S. and Goto, Y, *Nature Struct. Biol.*, 1996, **3**, 868–873.
216. Viguera, A. R., Martinez, J. C., Filiminov, V. V., Màteo, P. L. and Serrano, L., *Biochemistry*, 1994, **33**, 2142–2150.
217. Sosnick, T. R., Mayne, L. and Englander, S. W., *Proteins*, 1996, **24**, 413–426.
218. Zwanzig, R., *Proc. Natl. Acad. Sci. USA*, 1997, **94**, 148–150.
219. Schreiber, G. and Fersht, A. R., *Biochemistry*, 1993, **32**, 11195–11203.

ACKNOWLEDGEMENTS. We thank Dr M. K. Mathew for critical comments. This work was funded by the Tata Institute of Fundamental Research and by the Department of Biotechnology, New Delhi.

Received 16 December 1996; revised accepted 24 January 1997