

Invited Paper

## Protein Folding and $\beta$ -Sheet Proteins

Thallampuranam Krishnaswamy Suresh Kumar<sup>at</sup>, Thirunavukkarasu Sivaraman<sup>at#</sup>,  
 Dharmaraj Samuel<sup>a</sup>, Sampath Srisailam<sup>a</sup>, Gopal Ganesh<sup>a</sup>, Hui-chu Hsieh<sup>a</sup> ( ),  
 Kuo-wei Hung<sup>a</sup> ( ), Ho-Jen Peng<sup>b</sup> ( ), Meng-Chiao Ho<sup>a</sup> ( ),  
 Alphonse Ignatius Arunkumar<sup>a</sup> and Chin Yu<sup>at\*</sup> ( )

<sup>a</sup>Department of Chemistry, National Tsing Hua University, Hsinchu, Taiwan, R.O.C.

<sup>b</sup>Department of Medical Research and Education, Veteran Hospital-Taipei, Taipei, Taiwan, R.O.C.

The aim of this comprehensive review is to critically evaluate the progress in research in the area of protein folding. In the first section, we discuss the various models proposed to explain the protein folding paradox. In the succeeding section of the review, a detailed account of the developments in our understanding of the folding pathways of  $\beta$ -sheet proteins is provided.

The foundation for protein folding kinetics was laid about thirty five years ago. Christian Anfinsen showed that protein can fold reversibly and the native structures of proteins are thermodynamically stable states representing the global minima of their accessible free energies.<sup>1,2</sup> For most proteins, the information for folding is contained in the amino acid sequence.<sup>3-6</sup> However, a protein sequence needs to meet two criteria during folding – one thermodynamic and the other kinetic.<sup>7-14</sup> The thermodynamic requirement is that the protein molecule needs to adopt a unique folded conformation (native state), which is stable under physiological conditions.<sup>15-23</sup> The kinetic requirement is that the protein should fold back from its unfolded state to its native conformation within a reasonable time frame.<sup>24-34</sup> In principle, a polypeptide chain (in its unfolded state) can adopt numerous conformations.<sup>35-37</sup> Hence, if the protein needs to sample out all the possible conformations to decide on its global minimum, then for a protein consisting of 100 amino acids, it would take the age of earth (about  $10^{29}$  years if one assumes that only  $10^{-11}$  s is required to convert from one conformation to the other) to fold into its native conformation.<sup>38-40</sup> This puzzle is the crux of the protein folding problem and it is popularly called the Levinthal paradox.<sup>41</sup> Levinthal postulated that the thermodynamic and kinetic controls are the two mutually exclusive options for a folding molecule. Thermodynamic control meant that the folding molecule reaches its global minimum in energy and thereby folding is pathway independent. Due to the extensive search involved in this process, it takes a long time for a protein to fold under thermodynamic control. The kinetic control meant that folding occurs on a physiological time scale because it is pathway dependent.<sup>42-45</sup> Under the kinetic control,

the final structure of the protein could differ depending on the denaturant conditions from which folding is initiated.<sup>46-54</sup> Formulation of the Levinthal paradox led to the search for folding pathways.

The search for folding intermediates began in 1971. Tanford and co-workers reported stable intermediates along the equilibrium-unfolding pathway of hen egg white lysozyme.<sup>55-60</sup> Identification of equilibrium intermediates in bovine pancreatic ribonuclease consolidated the notion that protein-folding proceeds through definite structural intermediates.<sup>61-64</sup> These reports triggered the large modern enterprise of protein folding kinetics experiments which includes the studies on the slowing of folding by proline isomerization and use of thiol reagents to trap and characterize various disulfide intermediates.<sup>65-69</sup>

Many phenomenological models describe the events occurring during protein folding. These models generally focus on the Levinthal paradox and introduce ways in which only a very small fraction of the total number of conformers participate in folding from the unfolded state to the native state. Herein, we describe some of the popular and classical protein folding models.

### Diffusion-Collision Model

This model was postulated by Karplus and Weaver.<sup>70-74</sup> The basic tenets of this model stem from one of Anfinsen's proposal to describe the spontaneous folding of staphylococcus nuclease in solution. He proposed that portions of a protein chain could "flicker" in and out of their native framework and serve as nucleation centers that coalesce through non-covalent interactions to give rise to the native structure (Fig.

Present address: <sup>+</sup> Institute of Chemistry, Academia Sinica, Nankang, Taipei, Taiwan, R.O.C.

<sup>#</sup> Department of Biochemistry, College of Medicine, University of Iowa, Iowa City, IA 52242, U.S.A.

1). The diffusion collision model views a protein molecule to be divided into several parts (microdomains), each short enough for all conformational alternatives to be searched rapidly. This obviously means that the microdomains are so small that they are unlikely to be independently stable. Two or more such microdomains need to diffuse together, and collide in order to coalesce into a structural entity that is stable (Fig. 1). The process of protein folding is envisaged to involve numerous such collision coalescence steps to generate the native structure. These steps are ought to follow a unique order to yield a single structure for the protein.

### Framework Model

The framework model proposed by Kim and Baldwin has elements in common with the Diffusion-Collision model.<sup>75-77</sup> It is assumed that correctly hydrogen bonded secondary structure is formed prior to the formulation of the tertiary structure (Fig. 2). According to the Framework model, it is envisaged that folding is a hierarchical process in which simple structures are formed first which in turn act to give more complex structures.<sup>78-81</sup> The occurrence of the stable, partially structured intermediates called the “molten globule” states (to

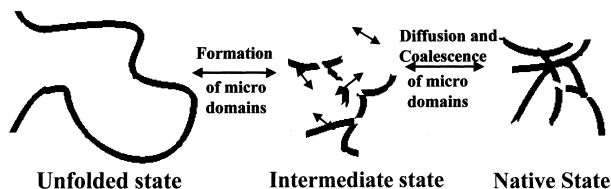


Fig. 1. Cartoon depicting the Diffusion-collision model. Upon initiation of folding from the unfolded state several microdomains are postulated to be nucleated. These microdomains are not independently stable and hence they are believed to diffuse together by a collision to a stable and unique three-dimensional structure.

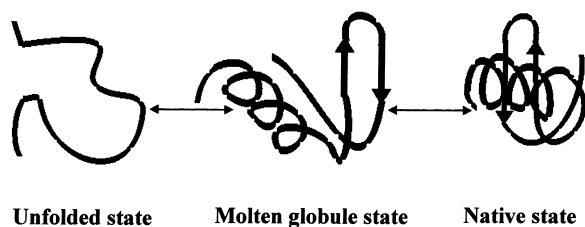


Fig. 2. Representation of the various steps in the Framework model of folding. Proteins are proposed to fold through intermediates that possess distinct secondary structural framework. Complete tertiary structural contacts are established subsequently to form the native, biologically active state.

be discussed later) in the equilibrium and kinetic folding pathways of protein validates the framework model.<sup>82-89</sup>

### Jigsaw Puzzle Model

Harrison and Durbin proposed a conceptual description of protein folding.<sup>90</sup> Proteins are viewed to fold by a number of different, parallel pathways instead of by a single definite sequence of events. This aspect would ensure folding to be robust to mutations that would not drastically affect the native structure. The Jigsaw model shows strong resemblance to the diffusion-collision model. It is argued that if all the elementary microdomains have similar properties then the multi-microdomain intermediates have similar folding and unfolding rates. Given this situation, even a small protein with only a few microdomains will have multiple alternative pathways (Fig. 3). However, as the kinetics of folding is sensitive to changes in free energies, only a small number of pathways are expected to dominate with the remains only turning out to be redundant.

### Nucleation Model

This model was originally proposed by Wetlaufer.<sup>91</sup> The model postulates that a portion of a polypeptide chain, unstable by itself, serves as the nucleus for chain propagation leading to the formation of the native structure. This model requires the existence of a nucleation unit small enough for a random search and once this basic unit has the native structure, a sequential and independent folding of subsequent amino acid residues is possible. The original “nucleation model” was further modified by Go and co-workers in a Growth-merge model.<sup>92</sup> The nucleation model was extrapolated to a “cluster” model by Kanehisa and Tsong.<sup>93,94</sup> At an early stage of the folding process, the polypeptide chain is assumed to be comprised of several locally ordered regions (called “clusters”) connected by random coil chain portions. Thus, in the denatured state, the size and distribution of the “clusters” are postulated to be dependent on the external con-

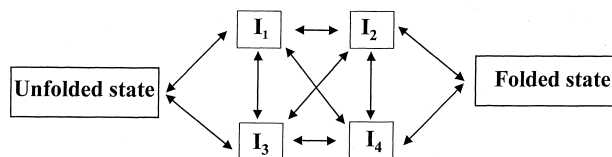


Fig. 3. Schematic representation of the folding events as perceived by the Jigsaw model. Folding is believed to proceed through multiple, mutually reversible pathways. The folding iteration culminates with the formation of an energy minimum native state.  $I_1$ ,  $I_2$ ,  $I_3$  and  $I_4$  are the various intermediates in the folding/unfolding path way(s).

ditions used. During the folding process these small “clusters” associate to yield the native state of the protein. The native state is perceived as a maximum size cluster with possible fluctuation sites. The “cluster” hypothesis has many elements in common with diffusion-collision model. However, an important difference may arise as to what governs the kinetics of the folding process. Is it growth of a single embryo (micro domain) or the coming together (diffusion-collision-coalescence) of two or more of these elements? Everything else being same, the nearest neighbor micro domains have higher probability to coalesce first, either by “growth” or by “collision”. Rapid folding of small proteins has been rationalized in terms of nucleation like mechanism.<sup>95-103</sup> Extensive mutational analysis on chymotrypsin inhibition 2 (CI 2) has provided evidence for the nucleation collapse mechanism.<sup>43,104,105</sup> Fersht and co-workers have proposed that certain residues have most of the native secondary structure fully formed in the transition state (TS) as compared to others.<sup>104-106</sup> However, it is currently not possible to establish if the transition state (folding nucleus) is unique or non-specific. Elaborate protein-engineering experiments of Serrano and co-workers on  $\alpha$ -spectrin SH3 domain<sup>107,108</sup> and also the thermal unfolding simulations by Daggett and co-workers proved that, the nucleation sites are not specific.<sup>21</sup>

Nucleation-collapse mechanism also draws strong support from lattice models.<sup>100,109-112</sup> Using cubic lattice models of proteins, it has been proposed that for “highly optimized” sequences, the folding nucleus is specific. In other words, the formation of tertiary structural contacts between certain “hot” residues with unit probability is a necessary and sufficient condition for folding. Guo and Thirumalai using off-lattice models of a sequence showed that there are numerous delocalized nucleation sites with some more probable than others.<sup>100</sup>

One of the ways to decrease the effective number of conformations searched is to employ two-stage kinetics. The first stage being compactization of the polypeptide chain into structureless globule. The search for the native conformation takes place in the second stage among the compact globular conformations whose number is drastically less than the total number of possible conformations. The two-stage kinetics was also observed in computer simulations of protein folding.<sup>113</sup> It was proposed that compactization of the polypeptide chain in the burst phase eases the subsequent random search for a transition state.<sup>114-116</sup> However, this logic is not likely to be applicable to proteins with large chain lengths since in longer polypeptide chains, the number of compact conformations is still exponentially large. Compact intermediates have been observed in the burst phase of refolding of many proteins.<sup>117-123</sup> Hydrogen-deuterium exchange experiments have

indicated that the compact states (identified in the burst phase of folding) lack specific structure.<sup>124-132</sup> Khorasanizadeh et al.<sup>133</sup> studying the tryptophan-containing mutant of ubiquitin at 25 °C, showed that burst phase intermediate is a compaction with few specific contacts and some secondary structure and is possibly fluctuating without any specific localization.

Sosnick et al.<sup>134</sup> in a recent study questioned the generality of intermediate(s). They demonstrate that cytochrome C at pH 5.0 represents essentially a two-state kinetics wherein all observed properties become native like simultaneously. Jackson and Fersht, studying the folding kinetics of chymotrypsin inhibitor (CI 2), showed that the protein folds with no detectable intermediate(s).<sup>105</sup> Interestingly, kinetic studies of the folding behavior of ubiquitin revealed that the absence or presence of compact states in the early stages of folding depends on the final concentration of the denaturing agent after dilution.<sup>131,135</sup>

Gutin et al.<sup>136</sup> examined the relevance of the non-specific hydrophobic collapse, in the very early stages of folding using a cubic lattice model (Fig. 4). In this model, a protein chain is considered to be positioned on a cubic lattice. An amino acid is presented by structureless monomer occupying a lattice site and residues connected by a covalent bond occupy neighboring lattice sites. Their simulation studies showed that depending on the value of the average interaction between monomers, two different regimes of folding are possible. A two-stage kinetics is observed when strong interaction between monomers dominates.

Camacho and Thirumalai suggested a three-step solution to the Levinthal paradox.<sup>95,137</sup> In the first step, the polypeptide chain is proposed to undergo compactization. At

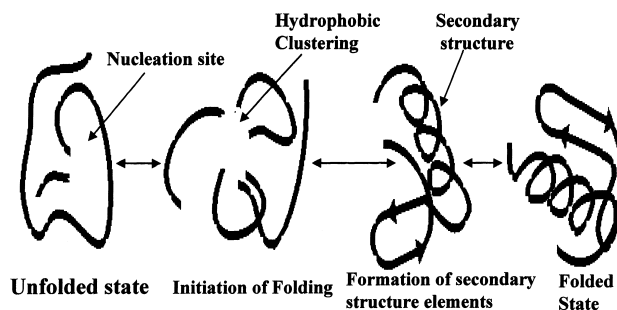


Fig. 4. Depiction of the hydrophobic-collapse version of the nucleation model. The earliest folding event is proposed to be the clustering of remote hydrophobic residues in the protein. Such clustering of the non-polar residues (in the initial phase of folding) is believed to decrease the conformational search during folding. Subsequent reorganization of the backbone is proposed to yield the native fold.

the second stage, the collapsed (or compact) chain is envisaged to fall into one of the minimum energy compact conformation. The native conformation is searched over a relatively small number of minimum energy compact conformations in the final stage. This hypothesis however ignores the fact that though the number of relevant conformations are decreased by chain compactization, energy barriers between them grow to an extent that interconversion between these fewer conformations slows down dramatically. In deed, several simplified protein models demonstrate that barriers between low energy misfolded states represent deep traps and escape from these is extremely slow. Thus, the proposal based on the step wise decrease of the number of available conformations may not provide an adequate picture of the folding kinetics as they actually replace entropic barriers by equally unsurmountable energetic ones.

### The New View of Protein Folding

The new view is based on the energy landscape perspective.<sup>16,22,138-145</sup> The new view recognizes that “folding pathways” are not the correct solution to the kinetic puzzle posed by Levinthal. The landscape perspective unambiguously explains the process of reaching a global minimum in free energy (satisfying Anfinsen’s experiments) and achieving it quickly (explaining Levinthal’s concerns) by multiple folding routes on funnel like energy landscape. The new view of folding sheds the conventional view of protein folding as a process wherein all polypeptide chains perform essentially the events in the same sequence, to reach the native state. The new view postulates that folding represents an ensemble average of process wherein is microscopically heterogeneous. Thus, each protein molecule could follow its own trajectory across the energy landscape to reach the global minimum (native state).

The energy landscapes (for protein folding/unfolding) figuratively could be represented by two lateral co-ordinates, (Fig. 5). These co-ordinates represent the degrees of freedom of the folding chain such as the torsion angles of the backbone and side chains. However, in principle a protein could have thousands of lateral co-ordinates. The vertical co-ordinates of the energy landscape represent the internal free energy of the polypeptide chain as it folds/unfolds. In practical terms, internal free energy represents the sum of all the intra-chain enthalpies and solvent interactions of a molecule in any given conformation. In an analogical terms, the kinetic process of folding/unfolding (as per the new view of folding) of a protein could be visualized as rolling a ball on a rough energy “terrain” (Fig. 5). Upon initiation of folding, the protein moulds its conformation in possible ways, which tend to decrease the energy. During its energy-decreasing endeavor, it is also buf-

feted into different conformation by Brownian motion. This aspect ensures that folding/unfolding steps against the energy barrier do occur but at very low frequency.

There are clear differences between the new models and classical models in capturing some of the molecular nature of proteins.<sup>141</sup> Unlike the phenomenological models, which treat only a few macroscopic-symbol states, the new models recognize the bond connectivity concept of a polymeric chain (such as in proteins) built from many monomeric units and protein compactness is only limited by excluded volume. These features permit an unbiased exploration(s) of the full ensemble of all conformations available to the polypeptide chain. In addition, the new models strikingly differ from the classical models in not assuming single or multiple exponential behaviour. These models also do not assume that folding/unfolding pathways proceed through the occurrence of identifiable intermediate macroscopic states. Unlike the classical models, the new models do not assume that macroscopic intermediate states (if they exist) are independent of environmental conditions such as the nature of denaturants, temperature, pH, etc. It should be recognized that new models have their own limitations. These models neglect atomic detail and often interpret folding based on shortened chains simplified energies and chain representations. In this context, it should be realized that modeling protein folding through energy funnels is just a conceptualization of the events and probably not truly applicable to any real protein. However, these modeling procedures could be useful tools to illustrate principles, designing new experiments, and posing new questions.

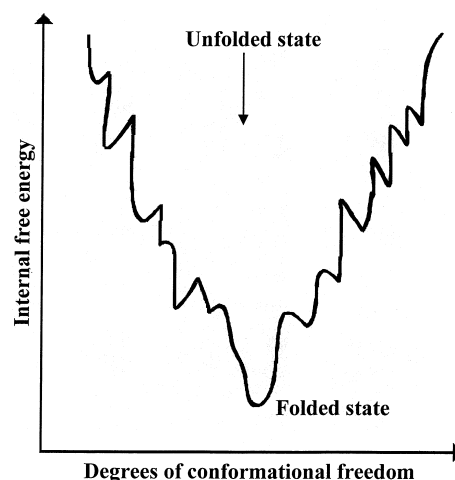
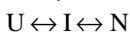


Fig. 5. Graphic representation of the protein folding landscape. Unfolded protein molecules are believed to mould their conformation in all possible ways, which tends to decrease the energy. Folding/unfolding steps against the energy barrier do occur but at very low frequency.

### On and Off Pathway Intermediates

For productive refolding/unfolding of proteins, it is important to ensure that the accumulated intermediate(s) are on the pathway leading to the native or denatured state(s). If the folding intermediate(s) are on-pathway then, the folding process could be elucidated by characterizing the intermediate(s) (Fig. 6). However, in the contrary, very little information could be gained by characterizing the structures of intermediate(s) which have drifted away from the folding/unfolding pathway(s).<sup>146</sup> Recently, kinetic intermediates have been characterized that have many structural features similar to the intermediate(s) which have been realized along the equilibrium unfolding pathway(s) of proteins.<sup>108,147-150</sup> These results imply that the intermediate(s) are on-pathway. Peng et al.<sup>36</sup> recently reported that the molten-globule intermediate of the  $\alpha$ -lactalbumin alpha domain, not only possesses a native secondary structure, but also has a native tertiary fold.<sup>36</sup> Similarly, Fersht and co-workers using the fractional change in free energy of the transition/native state produced by mutation ( $\phi$ ) values observed that in the barnase folding intermediate when plotted against residue number as those in the transition state.<sup>151-155</sup>

Formation of off-pathway intermediate(s) are well-documented in the literature.<sup>156</sup> The formation of incorrect disulfide pairing in the disulfide refolding pathway of proteins is a good example of the formation of off-pathway intermediates (Fig. 5, Ref. 157). Similarly, ligation of heme during the folding of cytochrome C by non-native side-chain, is another instance of accumulation of off-pathway intermediates during protein refolding. In general the on- and off-pathway models of protein folding in intermediates could be represented as, on-pathway model,



off-pathway model,

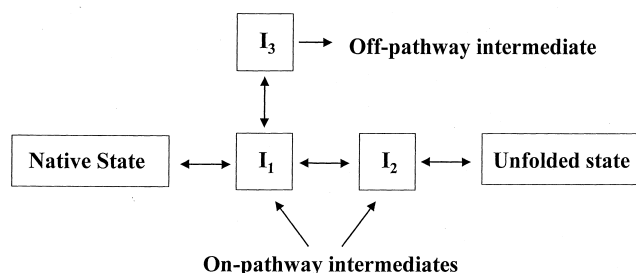


Fig. 6. Schematic representation of the 'on' and 'off' pathway in intermediates occurring during protein folding. 'Off' pathway intermediates are considered to be unproductive dead-end intermediates, which in many cases lead to aggregation of the protein.

wherein, I, U and N represent the intermediate, unfolded and native states of the protein. Recently, Laurents et al.<sup>158</sup> devised a pulse chase competition experiment to determine if a folding intermediate is on or off-pathway. They use a modified version of the traditional pulse-chase experiment in molecular biology to decide if the intermediate is "on" or "off" pathway.<sup>159</sup> The native protein with its amide protons labeled with  $^1\text{H}$  is compared when separate samples of  $^1\text{H}$  labeled intermediates (I) and unfolded protein are allowed to refold in  $\text{D}_2\text{O}$  at an appropriate pH to the native state under identical conditions. The pH is adjusted in such a way as to allow the protein in the unfolded state but not the off-pathway intermediate to rapidly exchange its backbone amide protons with the solvent. However, if the intermediate is on-pathway, then more  $^1\text{H}$  label is expected to be retained with the test sample starting from the intermediate than in the control sample starting with the unfolded state. Using this approach, Laurents et al.<sup>158</sup> demonstrate that the intermediate accumulated in low concentrations of guanidinium hydrochloride in RNase is a productive on-pathway intermediate. It should be noted that two criteria need to be met to satisfactorily employ the modified pulse experiments to determine whether the rapidly formed intermediate is 'on' or 'off'-pathway: (1) The kinetics of formation of the intermediate (I) must be measurable, (2) Simulation of the folding and exchange kinetics in step 2 ought to be made to decide if the test for an on-pathway intermediate is possible. On the whole, the most positive aspect of the modified pulse-chase experiments is that they do not give a false-positive result. The intermediate, "I" could be considered on-pathway, if a significant difference is found between the retained label in the test sample and the control.

### Hierarchical and Non-hierarchical Models of Protein Folding

In principle, folding pathway(s) of proteins could be classified into two broad categories – hierarchical and non-hierarchical. Hierarchical folding of proteins could be defined as a process folding starts with structure(s) which are of marginal stability and local in the sequence context.<sup>160</sup> It is presumed that these local structures interact resulting in intermediates with increasing degree(s) of structural complexity, which finally form the native conformation.<sup>161</sup> In the non-hierarchical process the tertiary structural interactions not only stabilize the local structures but also determine them (Fig. 7, and Refs. 88, 160). In short, in hierarchical folding, protein secondary structure(s) is determined by local sequence information. However, in non-hierarchical folding, non-local tertiary structural interactions dictate the nature of secondary structural interactions that form during protein folding.<sup>160</sup>

There are three lines of experimental evidence to indicate that protein folding is a hierarchical process. Firstly, many peptide fragments excised from proteins either possess or show a strong tendency to adopt the "native fold" even in the absence of long range interactions.<sup>160</sup> For example, the N-terminal helical 1-13 residue peptide in RNase (C-peptide) has properties expected from the parent protein structure, in which the helix extends from Thr3 to His12. Interestingly, the helical segment terminates at residue 12 even in the longer S-peptide (residues 1-20) fragment from RNase. This finding focussed attention on the local side-chain interactions that serve as helix-stop signals. The salt bridge between the charged side chains of Glu2 and Arg10 and the pseudo hydrogen bond between Phe8-His12 stand as good examples wherein, local side-chain interactions control the propagation of secondary structural elements in peptides/proteins.<sup>162</sup> The second line of evidence for the hierarchical folding stems from the experimental observation that helix-stop signals, which determine the boundaries of helices in proteins are in local sequences which surround each helix terminus rather than in residues that make tertiary structural interactions.<sup>163</sup> Analysis of 1316 protein helices by Creamer et al.<sup>162</sup> revealed that practically every helix has its own termination signal. Most of the helical segments in proteins have a distinct hydrophobic capping at the N- and C-termini and backbone hydrogen bonding is evident in nearly half of the helical segments analyzed. In general, the initial four NH groups and final four CO groups lack intrahelical hydrogen bonding partners thereby weakening the helix ends. These end-effects are significant and encompass more than two-thirds of the residues present in a  $\alpha$ -helix of an average length of 12 residues. In addition, the helix geometry at the helix N-terminus severely hinders solvent access to amide groups. These results clearly demonstrate that helix in proteins could be stabilized and/or terminated by hydrogen bonds between a side chain or main chain

group and backbone peptide groups at the helix ends. The third line of evidence for the hierarchical folding process comes from the characterization of folding intermediates.<sup>160,161</sup> Analysis of the structural interactions of the folding intermediates suggests that protein folding is a hierarchical process. Among the intermediates characterized to-date are those, which occur along the equilibrium unfolding pathway(s).<sup>164-168</sup> Using two-dimensional NMR hydrogen exchange in conjunction with quenched flow pulse labeling measurements of exchange, it was found that the kinetic and equilibrium intermediates are structurally equivalent.<sup>169-172</sup> These folding intermediates possess native secondary structure with the absence of persisting tertiary interactions. Interestingly, the native secondary structures range from partial to complete in these intermediates. For example, the acid form of Cyt C has all three major helices present in the native state, whereas the pH 4 intermediate of apomyoglobin has only three out of the eight helices present in the holomyoglobin.<sup>117,164</sup> Thus, these results wherein native like secondary structures are formed even though tertiary structural interactions are absent, represent the most convincing evidence for hierarchical folding.

The observation that segments of identical sequence could adopt different conformation(s) in different proteins provides a major challenge to the proponents of the hierarchical model. Minor and Kim devised an 11-residue sequence called the "Chameleon" sequence.<sup>173</sup> They introduced this sequence at two sites (A and B) in protein G. Protein G consists of a central helix (residues 23-25) surrounded by a four-stranded  $\beta$ -sheet. Site A comprises of 11 residues spanning residues 23 to 33. Site B, consisting of residues 42-52 overlaps a part of penultimate  $\beta$ -strand, a turn and part of another  $\beta$ -strand. When placed in site A, the 'Chameleon' sequence adopts helical conformation. Interestingly, the "Chameleon" sequence forms a strand turn strand conformation when situated at site B. Thus, it appears that the conformation of the "Chameleon" sequence is determined by its context within the total protein and not by local interactions.

$\beta$ -lactoglobulin, a protein isolated from milk source is yet another example which does not fit into the hierarchical model.<sup>161,174</sup>  $\beta$ -lactoglobulin is a 162 amino acid predominantly  $\beta$ -sheet protein containing two disulfide bridges. The protein also has one free cysteine residue. The secondary structural elements in the protein include nine antiparallel  $\beta$ -strands and a 11 residue helix. It has been shown that during the process of  $\beta$ -lactoglobulin refolding from a denaturant-induced unfolded state, the far UV ellipticity (219 nm) transiently exceeds the native intensity (the "overshoot" phenomenon). It is argued that the CD "overshoot" phenomenon is related to the accumulation of a  $\alpha$ -helical (non-native) intermediate

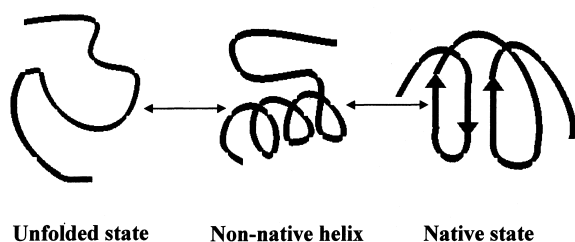


Fig. 7. Depiction of the non-hierarchical model of protein folding. Non-native secondary structural elements are formed in the early stages of folding of proteins. As folding proceeds, these non-native secondary structures reorganize and disappear to yield native secondary structural elements.

mediate.<sup>174</sup> The non-native helical intermediate formed during the burst phase of folding is believed to help in reducing the conformational search for the global minimum by decreasing the number of accessible conformations. As folding proceeds, the non-native  $\alpha$ -helical conformation is proposed to be transformed to the native  $\beta$ -sheet conformation, as this is more stable in terms of global free energy. It is worthwhile to note that the “over shoot” of the CD signal has been also observed in the burst phase of several other proteins. Chaffotte et al.<sup>174</sup> opined that the CD “over shoot” is an spectral artifact and stems from the far UV contributions of the aromatic amino acids and the disulfide bonds in the protein.<sup>175</sup> Sivaraman et al.,<sup>176</sup> recently studying the refolding of cardiotoxin in an analogue III, CTX III, from the venom of Taiwan Cobra (*Naja naja atra*) found that CD “over shoot” phenomenon is due to the asymmetrization of the disulfide bonds during the refolding of the protein. Despite the existent contradictions, the experimental evidence on  $\beta$ -lactoglobulin strongly supports that the folding of the protein ( $\beta$ -lactoglobulin) proceeds *via* a non-hierarchical process. In addition, the characterization of the structures of intermediates of several proteins such as Che Y and lysozyme suggests that non-native interactions but not non-native secondary structure(s) play roles in stabilizing the folding intermediates.<sup>175,177</sup>

Tendamistat and CTX III are two all  $\beta$ -sheet proteins whose unfolding/folding pathways could stand out as examples of non-hierarchical folding. Characterization of a partially folded equilibrium intermediate in tendamistat at pH 2-3 revealed induction of native-helical segments (25%) in 3-6 M trifluoroethanol.<sup>178</sup> The induced helical segments appear to be concentrated in regions corresponding to loops or random structure in the native state of the protein. Similarly, a stable intermediate has been characterized in the alcohol induced equilibrium-unfolding pathway of CTX III.<sup>179</sup> 25% of the backbone of the protein is found to exist in a helical conformation (by circular dichroism). Since, most of the native  $\beta$ -sheet interactions are present in the intermediates, it is believed that some of the residues in the loop or random coil region(s) in the native state of the protein transform into helix conformation.<sup>179</sup> It should be noted that existence of isolated, non-native secondary structure in structure-less regions (in the native state of the protein) does not *per se* validate the hierarchical mechanism of protein folding.<sup>160</sup> Only when the non-native structures observed in the intermediate structures interact productively with other local structures to yield higher order folding (in the intermediates), the folding of the protein could be considered non-hierarchical.<sup>180</sup>

### Transition States in Protein Folding

It is generally believed that folding pathway of a protein

will be solved, when the structures of all the stable, metastable and transition states of protein, “on” and “off” the pathway in intermediates are characterized both structurally and energetically.<sup>14</sup> The conventional definition of an intermediate in a chemical reaction is that it represents a structural entity which is at a minimum in a potential energy surface, and if any of the bonds are broken or reformed would result in a structure which is energetically less stable (Fig. 8). In contrast, a transition state in a reaction pathway is representative of a physical state wherein any change in bonds would lead to the formation of structures with increased stability. Thus, in terms of the three-dimensional energy surface, transition state in chemical reaction pathways is at a saddle point being at a maximum along the reaction co-ordinate. It should be noted that the transition states in simple chemical reactions and in protein folding pathways differ significantly. Only few bonds need to be broken or formed in the transition state of simple chemical reactions and in contrast many bonds have to be broken and made in the transition states of protein folding pathways. This subtle difference could account for the transition states in protein folding to be at very wide and long saddle points, with slight variations in structure leading to only small changes in energy.

The only option to characterize transition states is by inference from kinetic measurements on proteins wherein the structure – activity relationship is altered by site-specific mutations using protein-engineering methods.<sup>151,152</sup> The logic of employing protein-engineering methods to understand the structure of transition state(s) in protein folding is very simple.  $\phi$  analysis has been used to quantify the extent to which a given side chain stabilizes a transition state, relative to the extent to which it destabilizes the native state.<sup>151</sup> The  $\phi_f$  value for

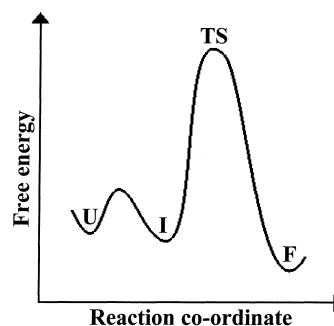


Fig. 8. Free-energy versus reaction co-ordinate plot for the folding of a protein *via* a folding intermediate. The reaction co-ordinate represents the path adopted as the reaction proceeds. The unfolded state (U), the intermediate (I) and the folded state (F) lie at the bottom of the potential energy wells, whereas the transition state (TS) is at a maximum.

a folding reaction could be defined by the equation,

$$\phi_f = \frac{[\Delta G^a(\text{WT}) - \Delta G^a(i)]}{[\Delta G_{\text{N-D}}(\text{WT}) - \Delta G_{\text{N-D}}(i)]}$$

“WT” and “i” represent the wild type and mutant respectively,  $\Delta G_{\text{N-D}}$  is the free energy difference between the wild type and the denatured protein and  $\Delta G^a$  represents the activation energy.

In the  $\phi_f$  analysis, most of mutations involve the replacement of residues by alanine because it has a minimal side chain. However, other replacements could also allow the estimation of the  $\phi_f$  value. In general, if the mutation destabilizes the folded structure (F) by  $\Delta\Delta G_{\text{F-U}}$  units of energy (which could be measured from the change in free energy of folding relative to the unfolded state (U)), then the free energy of a transition state (T) (measured relative to an unfolded state) changes by  $\Delta\Delta G_{\text{T-U}}$ . The  $\phi$  value is given by the expression,

$$\phi = \Delta\Delta G_{\text{T-U}} / \Delta\Delta G_{\text{F-U}}$$

This expression is analogous to the Bronsted equation frequently used in physical-organic chemistry.

$$\log k = \text{constant} + \beta \log K \quad (1)$$

wherein, “k” is the rate constant and “K” is the equilibrium constant for the reaction and  $\beta$  is the Bronsted co-efficient. The value of “ $\beta$ ” is a measure of the similarity of the structure of the transition state to the native state. Interestingly,  $\phi$  and  $\beta$  are the same at the two extremes zero and one when measured in the direction of bond making (folding).  $\phi$  (or  $\beta$ ) is equal to zero for the denatured state(s) and it assumes a value of “1” when the protein is in the native state.

$\phi$  can also assume fractional values.<sup>153,154</sup> The fractional  $\phi$  values could be due to weakened interactions in a single species (of the transition state) or due to the transition state being in a mixture of states, with some of the interactions completely formed and others fully disrupted. The mixture could arise from an equilibrium between the various structural states of the intermediate or due to the existence of parallel folding pathways giving rise to an ensemble of transition states. In addition, artifacts could arise due to distortion(s) in the structure of the folded molecule upon mutation. The structural distortions could introduce various reorganization energies (introduced by mutations) which do not cancel. However, analysis of the structural effects of a number of mutations at the same site would allow the detection of above mentioned artifacts in the  $\phi$  values. There is ideal gradation of the  $\phi$  values from 1.0 to close to ‘0’ upon mutation of residues in the middle to the edge of the  $\beta$  sheet in

barnase.<sup>151-154,181</sup>

The Bronsted equation could also be represented in free energy terms as

$$\Delta G^a = \text{constant} + \beta \Delta G_{\text{eq}}$$

“ $\Delta G^a$ ” is the activation energy and “ $\Delta G_{\text{eq}}$ ” is the free energy change at equilibrium. This equation could be successfully used to identify groups of residues that co-operate in forming the transition state. If the residues contribute to the stability of the transition state (T) to the same extent as the native state, then their Bronsted plot ( $\Delta G^a$  versus  $\Delta G_{\text{eq}}$ ) is expected to be linear. The position of the transition state could be gauged from the slope (m) values of the plot of  $\ln k$  (rate constant of folding/unfolding) versus the denaturant concentration. This plot is popularly called the “Chevron plot”. The Chevron plot for a two state folding is “V” shaped (Fig. 9). Initially the rate constant of folding ( $k_f$ ) decreases with the increase in the denaturant concentration (with a slope of  $m_f$ ) and beyond equilibrium the rate constant of unfolding ( $k_u$ ) shows a linear increase (with a slope of  $m_u$ ) with the concentration of the denaturant. The extent of folding in the transition state (T) is given by,

$$T = m_f / m_f - m_u$$

The position of the transition state is generally defined in terms of the amount of surface area buried upon D→T folding. Despite some inconsistencies, the transition state approximation is valid for protein folding as it provides a definite link between mutations and the consequent changes in the stability (of the transition state) and the folding and unfolding rates.

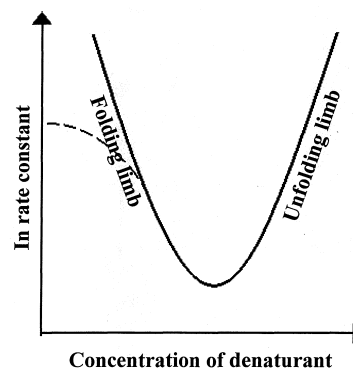


Fig. 9. A model of the Chevron plot. A perfect ‘V’ shaped Chevron curve indicates that the protein folds in a two-state (Native  $\leftrightarrow$  Unfolded) mechanism. A curvature (indicated by dotted lines) observed suggests the accumulation of folding intermediates during protein folding at low denaturant concentrations.

Recently, Munoz et al.<sup>31,145</sup> based on the results obtained by monitoring the thermal unfolding of an isolated  $\beta$  hairpin examined the transition state approximation. The data obtained on the unfolding kinetics was fitted to statistical mechanical model, which could specify the entire range of partly folded species obtained as a function of temperature from the time at which the unfolding of the peptide starts. It is found that 99% of the ensemble of the transition state species follow a minimal free energy path.<sup>31</sup>

Baldwin and Rose in a recent review article address an interesting question whether the transition states obtained in different folding reactions are similar.<sup>160</sup> They opine that the transition state barriers observed in folding pathways are low and broad, in contrast to those found in the ordinary chemical reactions which are high and sharply peaked. This opinion is supported by experimental results wherein single point mutations have been shown to cause large changes in the position of the transition state along the folding/unfolding pathway(s). Milla et al.<sup>12</sup> recently showed that point mutations in the *Arc* repressor causes a change from 0.92 to 0.69 in the position of the transition state in the direction of folding. In addition, the  $\phi$  values which yields information on the relative interaction strength of each residue shows a good correlation among structures of the transition states which occur along the folding pathway of barnase.<sup>152-154</sup>

To-date, the structures of the transition states have been characterized in limited number of proteins. Barnase, chymotrypsin inhibitor (CI2), *Arc* repressor, SH3 domains of *src* and  $\alpha$ -spectrin are the only few examples of proteins wherein the transition states have been structurally characterized along the folding pathway(s).<sup>160</sup> These proteins show a linear Bronsted plot for all residues implying that every residue affects the transition and the native states proportionally.

Transition state analysis based on characterization of mutants provides a snapshot picture of an intermediate state at one stage of the folding process. It can not give any information on the chronology of events leading to the formation of the transition state. Baldwin and Rose attempted to track the structural events to determine the hierarchy of the structural interactions in the transition state (in CI2) using the LINUS program<sup>182</sup> by suppressing the non-local interactions.<sup>160</sup> The results of this study revealed that large portion(s) of the structural framework is in-built in the amino acid sequence of the protein and it could be released in the absence of non-local interactions.

### Molten Globules in Protein Folding

Twenty years ago Tanford and co-workers<sup>56-60</sup> demonstrated that proteins undergo reversible unfolding in the presence of strong denaturants such as guanidinium hydrochloride

(GdnHCl). This led to a spurt in the search for intermediate state(s) along the unfolding/folding of protein.<sup>183</sup> Aune et al.<sup>183</sup> provided the most definitive evidence for the presence of residual structure in the acid and temperature denatured proteins. However, the concept of presence of intermediates with residual structures was contested by Privalov et al.<sup>184-186</sup> It was demonstrated that the changes of enthalpy, entropy and heat capacity upon unfolding are mostly the same for pH, temperature and GdnHCl induced unfolding. Based on these results it was argued that the pH and temperature unfolded proteins do not possess residual structure similar to the GdnHCl and urea denatured proteins. Privalov and Makhatadze<sup>186</sup> demonstrated the heat capacity changes during the acid and temperature induced unfolding of apomyoglobin, cytochrome C, RNase A, lysozyme are those expected for the completely unfolded polypeptide chains wherein the non-polar groups are exposed to water. Based on the thermal melting studies on small proteins Privalov suggested that proteins unfold and refold reversibly in an "all or none" fashion without stable and defined intermediates.<sup>184</sup> Kuwajima et al.<sup>77</sup> and Nozaki et al.<sup>55</sup> provided the first strong evidence for the existence of stable intermediates in the GdnHCl induced unfolding pathways of the bovine and human lactalbumin, respectively. Detailed characterization of the physical state of the protein molecules in these stable intermediate states using a variety of biophysical techniques revealed that the intermediate state(s) with similar structural features have been called the "molten globule" state(s).<sup>187-198</sup> Although the exact physical characteristics of the "molten globule" states are still a subject of intense debate, there has been some consensus on the physical definition of a "molten globule" state(s).<sup>199</sup> Proteins in the "molten globule" state are expected to possess the following physical characteristics.<sup>187</sup> (1) The protein molecule needs to be condensed with a Stokes radius equal or not greater than that of the folded protein, (2) should exhibit substantial native secondary structure with reduced stability of the constituent hydrogen bonds (3) should have lost most of the side-chain tertiary structural interactions stabilizing the native state. Small angle X-ray scattering, viscosity, sedimentation velocity and gel exclusion chromatography are some of the important techniques to probe the compactness of the "molten globule" state(s). Most proteins in their "molten globule states" possess a radius of gyration to be about 10% greater than that of the native proteins.<sup>187</sup> Assuming that if the molten globule structure(s) were empty with no water molecules contained in them, then the partial specific volume is expected to be greater than 50% of the native (or of the unfolded) protein molecule. However, since the increase in partial specific volume in the molten globule state (as compared to the native or unfolded state) is only about 5-10%, there is little doubt that the molten globule

states have a lot of water. The increase in the protein volume in the "molten globule" state is suggestive of water penetration inside the protein structure.

### Secondary Structure in Molten Globule States

There is a lot of debate concerning the extent of native secondary structure that is retained in the molten globule state. Circular dichroism spectra in the far UV region of a number of proteins suggest a high content of secondary structure in the molten globule states of various proteins.<sup>200-202</sup> In human carbonic anhydrase B, the negative molar ellipticity at 210 nm is nearly four times greater in the acid molten globule state than in the native state of the protein.<sup>200</sup> This aspect is not suggestive of a change in secondary structure because the far UV CD spectra are strongly influenced by contributions from the aromatic side chains. In general, the aromatic side chain groups lose their rigid, optically active environment in the molten globule-state leading to a decrease in their contribution to the far UV CD. Deconvolution of the far UV CD spectra into the contributions of the peptide and aromatic groups showed that the far UV CD in the molten globule-state differs from the native protein mainly due to contribution(s) of the aromatic groups.<sup>187,203-206</sup>

### Tertiary Structure in the Molten Globule State

The aromatic residues located firmly in an optically asymmetric environment in the native protein exhibit intense near UV CD. This aspect provides a good probe to monitor the tertiary packing of residues in the native state of the protein. In addition, fluorescence anisotropic measurements, spin echo, NMR techniques are the other techniques which have been used to study the state of tertiary structure packing in the molten globule state(s).<sup>187,207</sup> Plistyn and co-workers successfully studied the structural interactions using <sup>1</sup>H NMR spectra of the acid and temperature denatured forms of bovine  $\alpha$ -lactalbumin.<sup>208,209</sup> In principle the <sup>1</sup>H NMR spectra of a protein in the molten globule state is expected to be much simpler than in its native state. Most of the NMR resonances in upfield almost disappear in the molten globule state implying that the rigid architecture due to the packing of the aliphatic side chains and aromatic amino acids is drastically disrupted. Baum et al.<sup>199</sup> used these perturbed resonances for a tentative assignment of the NMR spectra of the protein ( $\alpha$ -lactalbumin) in the molten globule state. This study revealed that most of the native helical segments were retained in the molten globule state of  $\alpha$ -lactalbumin. Similarly, Baldwin and co-workers demonstrated the existence of at least three native helices in the molten globule-state of apomyoglobin.<sup>164</sup> Measurement of individual exchange rates for many protons showed that amide protons in helices A, G and H exchange 5 to 200 times

more slowly than in the unfolded state. The amide protons belonging to helix B exchange at 2 to 10 times more slowly than the unfolded states suggesting that the helix B is only partially folded or it is unstable. The amide protons exchange of the other residues in the protein was very fast and could not be traced. These observations led to the formulation that helices A, G and H in native state persist in the molten globule-state. Jeng et al.<sup>210</sup> characterized a molten globule state of cytochrome C at acid pH. Most portions of the three large native  $\alpha$ -helices remain protected from hydrogen-deuterium (H-D) exchange in the molten globule-state at acid pH and high ionic strength. However, the degree of protection of the residues in the three  $\alpha$ -helices was much lesser than observed in the native state. Interestingly, the NH groups that are involved in hydrogen bonding in the  $\beta$ -turns and in the tertiary structure of the protein were not significantly protected. Dobson and co-workers used a direct approach to study the tertiary fold of the molten globule-state by 2D-NMR spectroscopy.<sup>191-193</sup> Although the NMR spectra of the molten globule was very poorly resolved to be assigned by conventional methods, Dobson and co-workers used the magnetization transfer technique to achieve partial assignment.<sup>83</sup> They could successfully correlate the strongly perturbed aromatic resonance of  $\alpha$ -lactalbumin (from guinea pig) in the acid molten globule to those in the native state. This study concluded that cluster of aromatic groups also exists in the molten globule state. Recently, Dobson and co-workers elucidated the structure of a molten globule-like state of lysozyme in the alcohol-unfolding pathway.<sup>190,193</sup> Using a variety of experimental NMR strategies, they could demonstrate the structure of a partially structured state in 2,2,2-trifluoroethanol. This study revealed that most of the native secondary structural interactions characterizing the native state are intact in the alcohol-induced partially structured state of lysozyme.

Recent developments in NMR spectroscopy indicate that a detailed structural characterization of unfolded or partially folded proteins is within our grasp. A notable element in recent advances has been the availability of proteins enriched in <sup>15</sup>N and <sup>13</sup>C isotopes.<sup>193</sup> Availability of isotope labeled proteins enables the use of multi-dimensional heteronuclear NMR techniques to substantially enhance the resolution of the otherwise highly degenerate spectra characteristic of the non-native states. Recently, unfolded states of proteins in high concentrations of denaturing agents such as guanidinium hydrochloride and urea have been successfully characterized. This feature renders assignment and analysis of the NMR spectra possible. Neri et al.<sup>189</sup> recently structurally characterized the urea denatured 434-repressor protein. Similarly, Logan et al.<sup>211</sup> structurally characterized the FK506 binding protein unfolded in urea and guanidinium hydrochloride. Using

heteronuclear NMR, Shortle and co-workers reported the presence of residual structure in the denatured state of an unusual mutant of staphylococcal nuclease.<sup>190</sup> In general, the results from the above mentioned studies confirm the previous assumptions that local clusters of hydrophobic residues exist, at least in equilibrium with less structured states, particularly in the vicinity of the aromatic residues.

Stable, partially structured states have also been characterized in other environment(s). Partially structured states have been realized by dissolving by the proteins in mixed organic/aqueous solvents such as TFE. TFE is long been known to develop and stabilize extensive secondary structure in the absence of persistent tertiary interactions.<sup>212,213</sup> Recently, a molten globule state has been characterized in ubiquitin in 60% methanol at pH 2.0.<sup>86,196</sup> In the molten globule-state, the protein exhibits a NMR spectrum with relatively sharp and thus permitted assignment of a number of resonances. It was shown that three  $\beta$ -strands present in the native protein are also present in the molten globule state, both in their positions in the polypeptide chain and their mutual positions in space. The lone helix also remains intact in the alcohol-induced molten globule-state.

### Mobility of the Side Chain

It is generally found that the fluctuations of side chains in the molten globule are considerably higher than that in the native state. Semisotnov et al. demonstrated that the spin-spin relaxation time  $T_2$  of methyl groups in the molten globule coincides with that of the unfolded state.<sup>214,215</sup> Wong and Hamlin determined that the molecular volumes of the molten globule and native states are similar and hence predicted that the main resonance difference in the relaxation times could be due to intra-molecular movements.<sup>216</sup> The similarity of the spin-spin relaxation times for the molten globule and for the unfolded states suggests that the intra-molecular movements for the methyl non-polar groups in the unfolded and in the molten globule states are practically the same.

Rodionova et al.<sup>206</sup> determined that the motion(s) of the aromatic side chains in the molten globule state obtained in the urea unfolding pathway of carbonic anhydrase and found that the intra-molecular movements of aromatic side chains are much more hindered in the molten globule state than in the unfolded state. In contrary, measurement of the polarization of luminescence of Trp residues in  $\alpha$ -lactalbumin and bovine carbonic anhydrase showed that the intramolecular mobilities of the indole rings are nearly as restricted as in the native and molten globule states, while the restriction in the unfolded state is much lesser.<sup>217</sup> However, the aliphatic side chains have unrestricted motion in the molten globule-state. Shakhnovich and Finkelstein predict that there is sufficient space in side the

molten globule-state for the "free" movements of small and symmetric aliphatic groups.<sup>218</sup>

The large-scale fluctuations in the molten globule state(s) could be gauged from the field-dependent broadening of individual resonance in the proton NMR spectroscopy.<sup>217</sup> In the molten globule states, the fluctuations in the proteins are especially apparent in the aromatic groups. This aspect is consistent with the inter-conversion of different local conformations of the molten globule at rates slower than  $\sim 10^3 \text{ sec}^{-1}$ .<sup>218</sup> Another interesting approach to study the large-scale motions is to monitor the hydrogen-deuterium exchange. Dolgikh et al.<sup>208,209</sup> showed that the amide proton exchange is much faster in the molten globule state than in the native states in  $\alpha$ -lactalbumin and bovine pancreatic trypsin inhibitor. Interestingly, Merril et al.<sup>219</sup> demonstrated that accessibility of a protein molecule to proteases also increases in the molten globule state. Kallenbach and co-workers also arrived at the same conclusion by studying the relative susceptibilities of cytochrome C in the native and molten globule states.<sup>220</sup>

### Stability of the Molten Globule State

Shakhnovich and Finkelstein proposed that the molten globule state is stabilized by liquid like interactions of non-polar groups.<sup>217</sup> It is felt that the marginal increase in the molecular volume in the molten globule is sufficient to destroy the tight packing of the side chains, which consequently leads to the decrease in the van der Waals attractions. Based on site-directed mutagenesis studies Hughson and Baldwin showed that increase in side-chain hydrophobicity stabilizes the molten globule-state of apomyoglobin considerably but the same mutations are found to destabilize the native state.<sup>164</sup> These results authenticate the suggestion that the molten globule state(s) of proteins are primarily stabilized by non-specific hydrophobic interactions while the tight packing is important for the native state.

Molten globule state(s) contain considerable amount of native secondary structural interactions and the enthalpy of helix-coil transition in water is approximately about  $\sim 1.0$  kcal/mol. The enthalpy of hydrophobic interactions is small at 20 °C.<sup>221,222</sup> However, this parameter (enthalpy) is found to increase with temperature and thus possibly could contribute significantly to the heat effects of protein unfolding. Pfiel et al.<sup>221</sup> reported the enthalpy of the molten globule state of  $\alpha$ -lactalbumin at 4 °C does not differ very much from the enthalpy of the unfolded state. In general, the data on the thermal unfolding of the molten globule state are quite controversial. Gast et al.<sup>223</sup> showed that  $\alpha$ -lactalbumin does not melt co-operatively upon heating. This could be due to the protein existing in molten globule states at low pH and high temperatures. Using the micro-calorimetric data on the acid form of

retinol binding protein, Bychkova et al.<sup>222</sup> showed that the protein melts co-operatively with a small change in enthalpy. Pfeil et al.,<sup>221</sup> found that cytochrome C exists in a molten globule state at low pH and high ionic strength and the molten globule state melts on heating in a fashion similar to the native protein. However, it is still not known as to what extent these proteins are unfolded.

### Physiological Significance of the Molten Globule State

In recent years, the role of the "molten globule" intermediate has been well established. Proteins are involved in numerous cellular processes involving transmembrane transport.<sup>224-226</sup> These processes warrant major conformational changes. Translocation of proteins after their synthesis on the ribosome occurs in two phases. The first phase involves a time-dependent adsorption to the membrane and the second an ATP dependent process involving specific protein components. Dihydrofolate reductase (DHFR) which is a mitochondrial protein has been chosen as model protein to understand protein membrane trafficking.<sup>225</sup> *In vitro* experiments have revealed that the first phase of protein transport is accelerated when the protein (DHFR) is presented to the assay (membrane vesicles) system in its urea-unfolded form. This aspect possibly suggests that adsorption of the protein to the membrane surface are more readily accessed in the partially folded/unfolded state than the native state. This aspect was confirmed by the fact that the adsorbed form of the protein is found to be more sensitive to protease digestion than the native state. Interestingly, addition of methotrexate, a substrate of DHFR stabilized the protein against denaturation and also inhibited the initial adsorption phase. Bychkova et al.<sup>222</sup> opined that the physical characteristics of the molten globule state strongly suggest that protein translocation across membrane occurs via the "molten globule" state.

Van der Goot et al.<sup>227</sup> using colicin A, a membrane bound protein involved in pore formation, demonstrated that protein membrane insertions also possibly involve the molten globule state. The pore-forming domain of this protein includes ten well-packed helices and two of these helices are buried in the non-polar interior of the domain. Interestingly, the structure of the pore forming domain contrasts the proposed membrane insertion intermediate wherein the protein has been turned inside out with the two non-polar helices partitioned such that one of the hydrophobic helices moves into a polar environment leaving the other helix in the membrane. Interestingly, colicin A has been shown to exist in a molten globule state at pH 2.0. Taking into account the effect of the negatively charged lipid surface on the local pH, it was estimated that a pK of 3.0 was optimal for membrane insertion. Based on the above results it was proposed that colicin A ex-

ists in a molten globule state in the membrane surface. Similarly, the study of Kagan et al.<sup>228</sup> revealed that tumor necrosis factor (TNF) exists in molten globule state upon binding to membranes.

Recently, there are numerous studies implicating the involvement of chaperones in protein folding *in vivo*. Depending on the precise role on the folding pathway, chaperones are shown to bind to the target folding proteins in their unfolded or partially folded states. For example, GroEL, which is a popular folding chaperone has been unambiguously shown to bind to the molten globule-like folding intermediates.<sup>229</sup> However, one of the interesting aspects of chaperone binding is the lack of specificity for type or species of the protein. It is generally believed that the surface of the molten globule must be reasonably flexible to be deformable to a fairly standard shape and polarity that is largely complementary to the chaperone-binding surface.

*In vitro* protein degradation has also been suggested to involve molten globule-like intermediates. The degradation of proteins in lysozymes or ATP dependent proteosomes (as in the ubiquitin dependent proteolytic pathway) is suggested to be facilitated by prior denaturation, which may be the transition to the molten globule state.<sup>230</sup>

### Importance of Understanding Folding/Unfolding Process in $\beta$ -sheet Proteins

The advent of a wide range of sensitive techniques has led to the dramatic increase in the scope to examine mechanisms of folding.<sup>14</sup> Fast mixing methods wherein the folding reaction is monitored by sensitive optical techniques have provided invaluable information on the transition state(s) occurring during protein folding.<sup>135</sup> These techniques not only provide important data on the structure of the folding intermediates but also help to monitor the folding event in milli seconds to nano seconds time scales. Further, the application of modern NMR techniques has facilitated the conformation(s) of partially folded or unfolded proteins in greater detail.<sup>117,164</sup> The introduction of the native state hydrogen-deuterium exchange has enabled to probe the structure of partially folded state that accumulates transiently from their folded state.<sup>126</sup>

It is important that information gained by studying the folding/unfolding process in various proteins are effectively used to derive the general rules that govern the folding process. Critical review of the proteins wherein the folding studies have been examined show that very little information exists on the folding process of proteins, which are predominantly  $\beta$ -sheet. It is possible that the folding properties of proteins belonging to this class ( $\beta$ -sheet proteins) may differ significantly from those of helical and mixed  $\alpha,\beta$ -proteins because of the interactions stabilizing the  $\beta$ -sheets in proteins

are mostly non-local in nature. Hence, it is important that this lacuna is filled to ensure the generality of the concepts of the protein folding game. In addition, understanding the folding of  $\beta$ -sheet proteins is important from a clinical point of view. Recently the etiology of many hereditary disorders has been traced to involved defects in the folding of key proteins/enzymes.<sup>231-235</sup> The most prominent among these are the Alzheimers, Huntington's and variety of prion diseases.<sup>231-234</sup> These disorders involve the errant formation of  $\beta$ -sheet structures. In the background, it is apparent that better understanding of the folding properties of all predominantly  $\beta$ -sheet proteins would enable to delineate the molecular basis and evolve therapeutic strategies to tackle the protein folding related diseases. To-date, the folding/unfolding pathways of over ten proteins belonging to the  $\beta$ -sheet class have been investigated. In this review, we attempt to bring out the features of the folding/unfolding of various  $\beta$ -sheet proteins.

### Tendamistat

Tendamistat is a small (74 amino acids), two-disulfide bonded, all  $\beta$ -sheet protein (Fig. 10). The protein contains three Xaa-Pro peptide bonds, which are all in the *trans* conformation in the native protein.<sup>236</sup> Refolding kinetics monitored by change(s) in tryptophan fluorescence revealed that the refolding process occurs in three phases.<sup>237</sup> Interrupted refolding experiments were conducted to understand which of the three fluorescence detected phases produces the native protein. The results of these experiments revealed that refolding of the protein to its native state occurs in two parallel refolding channels and the intermediate observed in the kinetic reaction does not produce the native protein.<sup>237</sup> Schonbrunner et al.<sup>238</sup> investigated the origin of the kinetic

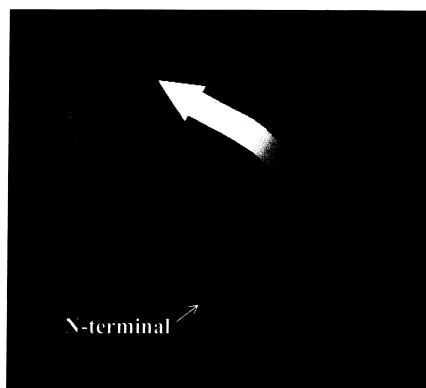


Fig. 10. MOLSCRIPT representation of the back bone folding of tendamistat. It is a 74-amino acid, all  $\beta$ -sheet protein with six  $\beta$ -strands running antiparallel to each other (not shown). The entire structure of the protein is held in position by two disulfide bonds.

heterogeneity observed in the refolding of tendamistat. They conclusively demonstrate that the two parallel pathways of refolding are due to the heterogeneity in the unfolded state caused by the slow *cis-trans* isomerization of the Xaa-Pro peptide bonds. The kinetic refolding of tendamistat is found to be two-state.<sup>238</sup> The slope of the logarithm of the unfolding and of the refolding rate constants versus the denaturant (GdnHCl) concentration is essentially linear over a broad range of the GdnHCl concentration. In addition, the free energy of stabilization estimated from the refolding and the unfolding rate constants according to the Eyring's equation also provide strong evidence for two-state folding. Schonbrunner et al.<sup>238</sup> also investigated if the rapid folding of tendamistat is due to the occurrence of a hydrophobic collapse during protein refolding. Clustering of hydrophobic residues are believed to restrict the conformational space of the polypeptide chain facilitating rapid refolding. Analysis of the results of the stringent fluorescence experiments showed the folding of the protein proceeds without rapid chain collapse.<sup>238</sup> The fluorescence properties of the solvent exposed tryptophan residue do not show a marked difference (that is expected in the event of a rapid hydrophobic chain collapse) during the refolding of the protein. In addition, the hydrophobic dye, ANS (1-anilino-8-naphthalene sulfonic acid), which is used to detect collapsed states in various proteins did not detect any solvent accessible hydrophobic surfaces during the refolding of the protein (tendamistat). It appears that although the conformational freedom in the protein during refolding is restricted due to the occurrence of the two-disulfide bonds in the protein, chain collapse does not seem to occur prior to formation of native tendamistat.<sup>238</sup> In essence, the study of Schonbrunner et al.<sup>238</sup> demonstrates that rapid chain collapse is not an essential step in protein folding. The influence of disulfide bonds on the refolding kinetics of proteins has been a subject of intense debate. It is believed that presence of disulfide bonds would yield kinetically trapped intermediate(s) during protein refolding. In the contrary, Camacho and Thirumalai<sup>95</sup> showed that disulfide bonds could enhance the rate of refolding by restricting the conformational space for the folding polypeptide chain. In this context, it is interesting to note that wild type tendamistat refolds rapidly in a two-state manner despite the presence of the disulfide bonds.<sup>237</sup>

Using a site-directed mutagenesis approach, Schonbrunner et al.<sup>178</sup> recently investigated the role performed correct tertiary interactions (such as disulfide bonds) on rapid two-state folding exhibited by tendamistat. As mentioned earlier, tendamistat contains two disulfide bonds between Cys11 and Cys27 and Cys45 and Cys73 which in-turn holds the  $\beta$ -strands tighter to form the triple-stranded  $\beta$ -sheet segment.

Schonbrunner et al.<sup>237</sup> prepared two single-disulfide

variants of the protein (tendamistat) namely, Cys111Ala/Cys27Ser and Cys45Ala and Cys73Ala, using the site-directed mutagenesis approach. Estimation of free energies of the unfolding and refolding using the Eyring equations showed that the two-state character of the folding/unfolding reaction is not changed by the disulfide bond replacements. Interestingly, the unfolding limb of the Cys45Ala/Cys73Ala mutant displayed a prominent kink (at higher denaturant concentration) which is observed in neither the wild type nor the Cys111Ala/Cys27Ser mutant. The shallower slope in the Chevron plot at higher concentration of the denaturant probably reflects a change in the unfolding mechanism.

The 11-27 disulfide bond in tendamistat is a local tertiary contact connecting the ends of a  $\beta$ -hairpin. Model peptide studies reveal that this region of the protein has a strong tendency to form a hairpin loop even in the absence of disulfide bond.<sup>237</sup> In comparison to the wild type, large changes were observed in the unfolding rate constants in the Cys111Ala/Cys27Ser variant. Such large changes in the unfolding rate constant (observed in the disulfide mutant) are suggested to indicate that in reactions in the native protein at the locations of the cross-links are weakened by replacement of disulfide bonds. The refolding rate constants were found to be much less affected by the mutations than those of the unfolding reaction. Although, the disulfide bond (Cys111Ala/Cys27Ser) connects parts of the molecule which have to find each other in the transition state, the preformed disulfide bonds (in the wild type tendamistat) do not seem to have a major rate enhancing effect in the refolding process. The effect(s) of enthalpy/entropy compensation observed on activation parameter of the folding reaction was rationalized on the residual structure in the unfolded state introduced by the disulfide bonds.<sup>237</sup> The residual structure is proposed to lead to unfavorable entropy in the unfolded state and also to reduce structural interactions in the unfolded state, which appears to be enthalpically favorable. The findings of the studies essentially suggest that preformed interactions such as disulfide bonds would favor refolding by decreasing the loss of chain entropy but would also at the same time disfavor the folding chain to gain conformational enthalpy upon formation of the transition state, as most of the interactions are already formed in the unfolded state (due to the presence of disulfide bonds). Thus, this study demonstrates that in  $\beta$ -sheet proteins, preformed correct interactions will have little effect on the rate of protein refolding. In addition, careful kinetic analysis of the single disulfide variants of tendamistat revealed that formation of the  $\beta$ -hairpin structure between residues Val12 and Gly26 is the rate limiting step in the folding process of the protein.<sup>237,238</sup> Interestingly, even when the disulfide bond tethering the two ends of the  $\beta$ -hairpin are replaced, this region (res-

idues spanning Val12 to Gly26) are found to form a compact and energetically very favorable structure in the transition state.

Thomas Kiefhaber and co-workers also investigated the effect(s) of a heliogenic solvent such as 2,2,2-trifluoroethanol (TFE) on the structure of tendamistat.<sup>238</sup> None of the residues in the protein exist in a  $\alpha$ -helical conformation in the native state. At low concentrations of TFE the protein is found to lose its defined tertiary interactions in a cooperative manner leading to the formation of partially structured state. The loss in the tertiary structural interactions are indicated by the disappearance of the near UV CD bands and decreased chemical shifts dispersion in the resonances of the 1D NMR spectrum. Systematic hydrogen-deuterium (H-D) exchange and FTIR experiments showed that most of the native  $\beta$ -sheet elements are intact in the TFE-induced partially structured state.<sup>239</sup>  $\alpha$ -helical structure in the alcohol-induced partially structured state of tendamistat are contemplated to be located mainly in regions corresponding to loops or random structure in the native protein.<sup>239</sup> The presence of a subset of the native long range interactions and the loss of stable native interactions in tendamistat suggested that the TFE-induced partially structured state represents an early intermediate in the hierarchical folding pathway of the protein.

### Fibronectin Modules

Fibronectin type III (FN III) module is an independent folding subdomain belonging to the fibronectin type III superfamily.<sup>240</sup> FN III domain derived from human tenascin is 91 amino acids long and the secondary structural elements in the protein are primarily  $\beta$ -sheets. The structure of FN III consists of seven  $\beta$  strands arranged into two  $\beta$  sheets, one of four and one of three strands (Fig. 11, Refs. 240, 241). The FN III domain is characterized by the absence of disulfide bonds and the presence of single highly conserved tryptophan residue. Interestingly, the protein contains eight proline residues. Plaxco et al.<sup>242-244</sup> recently monitored the refolding kinetics of this protein. The protein (FN III) despite the presence of nine proline residues is found to refold rapidly within 1 second. The rapid refolding of the FN III domain (from its GdnHCl denatured state) was unexpected since proline *cis-trans* isomerization is known to result in a population of slowly folding molecules in the refolding kinetics of many proteins. The rapid refolding of the protein is explained on the basis of the relatively high thermodynamic stability of the FN III domain.<sup>243</sup> It is presumed that the overall stability of the domain correlates to the stability in regions of native conformation in the transient folding nucleation sites.

Plaxco et al.<sup>243,244</sup> recently compared the folding kinetics and thermodynamics of two homologous fibronectin type

III modules (ninth 9 FN III and tenth 10 FN III modules). These two modules share identical topologies and core structures. These two FN III modules share less than 30% amino acid sequence identity. However, both the FN III (9FN III and 10FN III) modules comprise of seven  $\beta$ -strands arranged in two topologically complex sheets.<sup>240,241-245</sup> 9FN III contains seven proline residues as compared to 10FN III, which possesses eight proline residues. Equilibrium chemical denaturation experiments revealed that the 9FN III module is appreciably more stable than its homologue (10FN III). The relative difference(s) in the free energies of unfolding of the 9FN III and the 10FN III modules have been estimated to be 1.2 kcal mol<sup>-1</sup> and 6.1 kcal mol<sup>-1</sup>, respectively. However, no clear explanation is yet available to account for the differences in thermodynamic stability of these two homologous FN III modules. Both the FN III modules have been shown to fold slowly in a two-state manner. It was ensured that the slow folding was not due to either accumulation of non-productive intermediate or the occurrence of slow folding on-pathway intermediate(s). Kinetic analysis of the stopped-flow fluorescence and CD data revealed that the refolding of 9FN III proceeds in two co-operative parallel folding pathways. As the individual phases of the biphasic refolding pathway of the protein (9FN III) remained unchanged over a wide range of the denaturant concentration, it is presumed that the parallel folding pathways are not due to occurrence of different folding intermediates but due to heterogeneity of the denatured state(s).<sup>242-244</sup> 10FN III also shows a similar refolding pattern. Interestingly the intrinsic refolding rates of 9FN III and 10FN III differ by more than three orders of magnitude. 10FN III folds faster than the 9FN III module. This situation is in marked contrast

to other homologous proteins, which tend to share quantitatively similar folding pathways and have relatively similar refolding rates. Plaxco et al.<sup>243,244</sup> propose that with the high content of proline residues in the FN III (9FN III and 10FN III) modules, that at least 50 to 60% of the unfolded protein molecules could be predicted to contain one proline residue in the incorrect configuration. This aspect is reflected in the 67% amplitude of the slow phase of the refolding of 9FN III.<sup>243</sup> Interestingly, the 10FN III module does not exhibit a detectable proline isomerization-like phase. It is suggested that the absence of the slow phase could be due to the fact that the protein folds into a stable, native-like conformation, which is yet indistinguishable from the native protein.<sup>243</sup> 33% of the denatured 9FN III module which folds in a proline independent pathway were found to fold at rates 100 times slower than the 10FN III module.<sup>243,244</sup> The vast difference in the refolding rates is attributed to the relative thermodynamic stability of the two modules. Under conditions wherein the intrinsic stability of 9FN III and 10FN III are similar, the refolding rate constants of the protein modules are also similar.

### Carbonic Anhydrase

Carbonic anhydrase protein isolated from a variety of sources has a molecular weight of about 30 kDa. The enzyme consists of single domain and the secondary structural elements include a short helix and 10  $\beta$ -strands that divide the molecule into two halves (Fig. 12, Ref. 242). Each half contains a hydrophobic cluster. The hydrophobic cluster in the N-terminal region consists of four aromatic residues, which have been successfully used to monitor various aspects of the folding/unfolding behavior of this protein. In addition, the protein lacks disulfide bridges hence has been used as a popu-

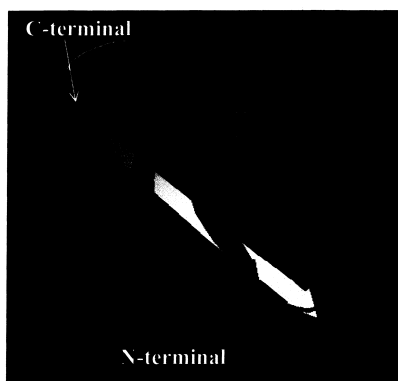


Fig. 11. Backbone folding of the Fibronectin type II domain (the tenth module from human fibronectin). The protein consists of seven  $\beta$ -strands arranged in two topologically complex sheets. The protein is characterized by high content of proline residues.

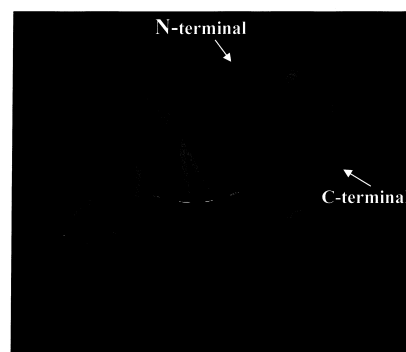


Fig. 12. Structure of carbonic anhydrase (human). The secondary structure of the protein consists of 10  $\beta$ -strands and a short helix. The  $\beta$ -strands divide the protein molecules into two halves with each containing a hydrophobic cluster. The protein in its active state is co-ordinated with zinc.

lar model for understanding the protein-folding problem. Carbonic anhydrase is among the first few proteins, wherein partially structured intermediates such as the 'molten globule' were been characterized.<sup>187</sup> Martensson and Jonsson et al.<sup>245,246</sup> using a variety of biophysical techniques showed that human carbonic anhydrase forms a stable and compact folding intermediate at a moderate concentration of guanidinium hydrochloride. They introduced single cysteine residues in various parts of the  $\beta$ -structure and monitored the stability and compactness of the area encompassing each cysteine residue by chemical labeling. Their study demonstrated that the folding intermediate of human carbonic anhydrase (in GdnHCl) has an ordered secondary structure in the central portion of the  $\beta$ -sheet. The peripheral part of the  $\beta$ -sheet structure was shown to be less ordered. Interestingly, the large hydrophobic cluster located between the central  $\beta$ -sheet core and the secondary structural elements on the surface was found to be unperturbed even at a high denaturant concentration. Svensson et al.<sup>247</sup> further characterized the structural properties of the 'molten globule' in intermediate using fluorescence and EPR measurements. Of the 10  $\beta$ -strands, strands 3 to 7 were shown to be unperturbed. In contrast, the peripheral  $\beta$ -strands 1, 2 and 9 were observed to be completely disrupted in the intermediate.<sup>247</sup> The hydrophobic core comprising  $\beta$ -strands 3-5, was remarkably stable even at 5 M GdnHCl. The stability of the hydrophobic cluster is found to increase toward the center. Oversky and Ptitsyn investigating the GdnHCl-induced unfolding of bovine carbonic anhydrase by size-exclusion chromatography in conjunction with optical methods showed that the protein unfolds at low temperature via two intermediates - the 'molten globule' states.<sup>248</sup> In the 'pre-molten globule' state the protein binds weakly to 8-anilino naphthalene-1-sulfonate and its hydrodynamic volume is two-fold larger than the native protein. The 'pre-molten globule' state is postulated to share a number of properties with the burst phase kinetic intermediate(s) occurring in the very early stages of protein refolding. The refolding of human carbonic anhydrase from its GdnHCl denatured state was monitored by near UV CD measurements using various tryptophan mutants (of the protein). This way the development of asymmetric environments around specific tryptophan residue during refolding was probed.<sup>249</sup> The N-terminal domain (residues 1-25) was found to fold slower than the major domain (residues 26-260) which contains the active site of the protein molecule. Thus, an essentially native structure of the major domain appears to be template for the correct folding of the N-terminal portion of the carbonic anhydrase molecule.<sup>249</sup> In addition, the tryptophan residue located in the hydrophobic cluster comprising the  $\beta$ -strands, 3-5, is observed to assume native-like asymmetric environment during the burst phase of refolding implying

that the hydrophobic cluster functions as a folding initiation site in the protein.<sup>249</sup> Refolding studies on human carbonic anhydrase have indicated that a kinetic intermediate that is observed after approximately 1 second of refolding under native conditions has some properties in common with the equilibrium intermediate observed in the GdnHCl induced unfolding pathway of the protein.<sup>250</sup>

### Cold Shock Protein B

Cold shock protein B (CspB) is a small molecular weight (~7.5 kDa), all  $\beta$ -sheet protein isolated from *Bacillus subtilis*.<sup>251</sup> The secondary structural element(s) in the protein include a single five-stranded  $\beta$ -barrel with no helical segments (Fig. 13, Ref. 252). Equilibrium unfolding studies on CspB indicate the absence of intermediate along its unfolding pathway. Interestingly this protein does not contain disulfide bonds and is not influenced by post-translational modifications or by tight binding to cofactor(s).<sup>253</sup> The thermodynamic stability of this protein is low but the unfolding and refolding of (CspB) are extremely rapid process.<sup>253</sup> The time constant for refolding is 1.5 ms. At the transition midpoint (wherein the folding and unfolding of a protein is usually slowest), CspB folds extremely rapid and the equilibrium between native and unfolded states is established in less than 100 ms.<sup>254</sup> The unfolding and refolding reactions of CspB are described by monoexponential process, and identical rate constant values are realized for refolding and unfolding. Complete amplitude change occurs during the single exponential unfolding/refolding process of this reaction indicating the absence of burst phase in intermediates. In addition the Chevron plots under varying denaturant concentration(s) shows a 'V'-shaped dependence on denaturant concentration with a minimum in the

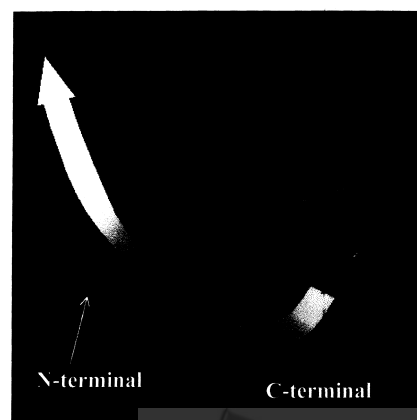


Fig. 13. Representation of the overall polypeptide folding of cold shock protein B. It is a ~7.5 kDa, five stranded  $\beta$ -barrel protein. The protein lacks disulfide bonds.

middle of the folding transition and an increase in the slopes of unfolding/refolding when moving away from transition midpoint. Interestingly, CspB has the most highly structured transition state of any protein studied to date ( $\alpha$ -volume of 0.86). In accordance with the high  $\alpha$ -value, 90% of the change between the unfolded and folded states occurs between the unfolded and transition states of the protein.<sup>254,255</sup> Folding studies on CspB clearly indicate that there is no correlation between the stability of the protein and its rate of folding. CspB is only marginally stable, but both the refolding under native conditions and the equilibration of the folded and unfolded states in the transition region are highly rapid reactions. Schlinder et al.<sup>253,254</sup> have explained the extremely rapid folding of CspB on the basis of Go's consensus principle. According to this proposal, folding could be extraordinarily rapid, if the folding energy gained during refolding process is not used to stabilize non-native interactions. It is proposed that a high consensus could be easily maintained when all potential folding intermediates are unstable and thus avoiding kinetic traps. Schlinder et al.<sup>253</sup> opine that the rapid folding of CspB could be attributed to the lack of intermediate(s) in their refolding reaction.

Schlinder and Schmid also investigated the folding kinetics of CspB as a function of urea concentration at various temperatures.<sup>253</sup> Interestingly, under all these conditions, the folding of the protein follows a two-state mechanism. The close adherence of the protein to the two-state mechanism encouraged these authors to employ the transition state theory for analyzing the observed kinetics and estimate the activation parameters for unfolding and refolding. During refolding in the absence of urea, 90% of the change in  $\Delta C_p$  and 96% of the change in 'm' occur between the unfolded and the activated state of CspB.

Peri et al.<sup>256</sup> investigated the question whether stability and kinetics of a protein are interrelated. They compared the stability of two cold shock proteins with CspB.<sup>256,257</sup> In contrast to CspB, which is only marginally stable, the other two homologous cold shock proteins (CspT and CspH) possess increased conformational stabilities. It is found that both CspT and CspH show very fast two-state kinetics like CspB. Thus, it appears that there is no link between conformational stability and its folding rate. This conclusion is also supported by studies on tendamistat (an all  $\beta$ -sheet protein, described earlier) wherein a strict adherence to a two-state model was found and the kinetic 'm' values of unfolding were insensitive to various mutations.<sup>237,238</sup>

### Cellular Retinoic Acid-binding Protein 1

Cellular retinoic acid-binding protein 1 (CRABP1) is a ~16 kDa predominantly  $\beta$ -sheet protein. The secondary struc-

ture of the protein consists of 10-stranded  $\beta$ -clam structure and a small helix-turn-helix segment (Fig. 14, Ref. 258). There is a large central cavity inside the clam shell-like structure, which is filled with the solvent in the absence of the bound lipid. This protein is prototype of a large family of proteins that bind to hydrophobic ligands. Although the members of the lipid binding protein family have sequence homologies ranging from 20% to 80%, they conspicuously have similar three-dimensional structure.<sup>259</sup> CRABP1 is best in interesting spectroscopic properties, which would endear this protein to a "protein folder." The fluorescence emission of CRABP1 (in its native state) is quenched due to an energy transfer from tryptophan to retinoic acid.<sup>260</sup> Unfolding the protein results in a strong increase in the fluorescence intensity with a concomitant prominent red shift of the wavelength of maximum emission. Similarly, unfolding the protein also results in a large change in the CD signals in the near and far UV region.<sup>261</sup> These significant differences in the CD and fluorescence signals are potentially useful probes to monitor the folding/unfolding behavior of this protein. In addition, the three tryptophan residues are located at structurally distinct locations in the protein molecule and hence serve as regio-selective structural probes during protein folding/unfolding.<sup>261</sup>

Liu et al.<sup>261</sup> examined the conformational behavior of CRABP1 under various denaturing conditions. Although urea-induced unfolding process of the protein fits into the two-state model (Native  $\rightleftharpoons$  unfolded states). CRABP1 at low pH conditions is shown to exist in a partially structured state. In the acid-induced intermediate state, the helical content is substantially higher than in the native protein. At higher concentration of TFE 75% of the backbone of the protein is found to fold to a helical conformation. These results led Liu et al.<sup>261</sup> to speculate that the non-native structure could be involved in

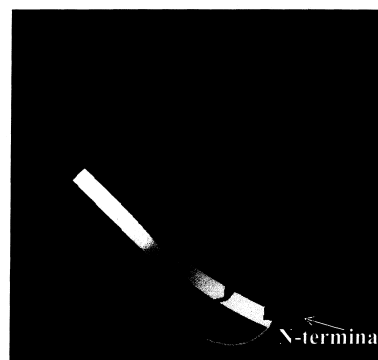


Fig. 14. Depiction of the structure of the cellular retinoic acid-binding protein. 10  $\beta$ -strands arranged in the form of a clam constitute the secondary structure of the protein. There is a short helix-turn-helix segment in the protein.

the folding mechanism of CRABP1. Interestingly, addition of sodium sulfate to the acid-induced intermediate state caused a conformational change resulting in the reversion to the predominantly  $\beta$ -sheet structure.

Clark et al.<sup>262</sup> evaluated the intrinsic tryptophan fluorescence as a probe for the structure and folding of CRABP1. Their study revealed that the three tryptophan residues which are not buried to the same degree in the protein, do not equivalently tolerate the Trp $\rightarrow$ Phe mutational changes. Interestingly, mutation of the fully buried tryptophan (Trp7) with Phe also does not affect the structure and stability of CRABP1. Stopped-flow fluorescence experiments on the wild type CRABP1 provides a strong indication of the presence of intermediate species. Intermediate(s) is formed during the very early stages of folding (<10 ms). The most noticeable aspect of the intermediate is its increased fluorescence emission intensity in the 320-330 nm region. These results indicate that a non-polar environment is formed around one or more tryptophan residue(s) prior to the formation of the native state. Further, detailed examination of the kinetics of the regainment of the intrinsic tryptophan fluorescence, indicates that the refolding time constants and amplitudes of the three tryptophan residues (in CRABP1) are similar implying that folding requires complete-chain involvement. The results described in this study suggest that the hydrophobic collapse event and not the formation of autonomous hairpin initiates folding of CRABP1.<sup>262</sup>

Clark et al.<sup>263</sup> recently compared the rates of formation of the central ligand binding cavity and the consolidation of the native hydrogen bonding network. The folding of CRABP1 monitored through tryptophan and ANS fluorescence involves at least four kinetic phases. The protein is mostly folded within a time span of 200 ms. The late kinetic phase is believed to be associated with the *cis/trans* isomerization. Kinetic hydrogen exchange NMR experiments reveal that formation of stable native hydrogen bonds in CRABP1 occurs in a concerted fashion. It is found that the protein binds to the ligand prior to the formation of the stable hydrogen-bonding network. ANS binding and quenched flow NMR data were obtained to suggest that the folding of CRABP1 occur through an early formation of a collapsed structure. Formation of this intermediate is followed by the appearance of a state with native-like three-dimensional structure, which includes the ligand-binding cavity. This structure does not contain the stable hydrogen-bonding network. It is proposed that the lack of specific tertiary structural packing cause hydrogen bonds to be temporary and weaker than in the native state. It appears that formation of the three-dimensional topology restricts the conformational space favoring the development of native contacts.

### Interleukin-1 $\beta$

Interleukin 1 $\beta$  is a member of a family of cellular mediators called as cytokines. The protein belongs to the all  $\beta$ -class and is made up of 12 antiparallel  $\beta$ -strands connected by turns, short or long loops (Fig. 15, Ref. 265). The long loop connecting the  $\beta$ -strands 3 and 4 is strange in that it contains a type II turn followed by 2-3 turns of a  $3_{10}$ -helix. Despite the presence of a  $3_{10}$ -helix, interleukin 1 $\beta$  could still be considered as an all  $\beta$ -sheet protein because the long loop is surface exposed and the short  $3_{10}$ -helix lies outside the case of the protein.<sup>266,267</sup> Interleukin 1 $\beta$  contains no disulfide bonds but has two solvent accessible sulfhydryl groups and four tyrosine residues, one which is adjacent to a single tryptophan residue at position 120 in the protein molecule.<sup>266</sup> Craig et al.<sup>268</sup> were among the first to investigate the conformational stability and folding of interleukin 1 $\beta$ . Interleukin 1 $\beta$  is a thermodynamically stable globular protein, with a  $\Delta G(\text{H}_2\text{O})$  of 29.4 kJ mol<sup>-1</sup>. The reversibility of the refolding is pH-dependent and the yield of the refolded protein decreases with increase in pH.<sup>268</sup> The protein completely unfolds at 2 M GdnHCl and above. The protein exhibit a high co-operativity under conditions where refolding is reversible. The yield of the refolded protein is close to zero if the protein is allowed to stand in concentrations of GdnHCl and the protein aggregates at higher concentrations of the denaturant.<sup>268</sup> Similarly, the reversibility of the protein from the 8 M urea-unfolded state (at pH 7.8) is poor. Refolding of the protein proceeds with a biphasic kinetics. The slow phase has been shown to be not associated with proline isomerization.<sup>268</sup> This phase has been proposed to be due to slow rearrangement from the compact collapsed, compact globular state (formed in the initial phase) to form the native state of the protein. Varley et al.<sup>269</sup> examined the kinetics of refolding of interleukin 1 $\beta$  in greater detail using nuclear magnetic reso-

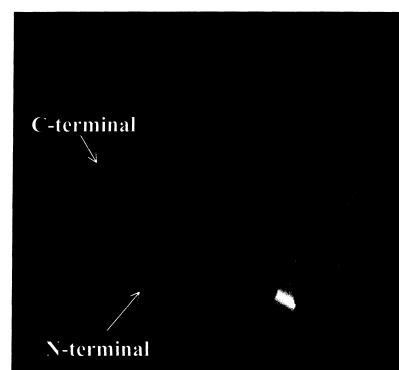


Fig. 15. The  $\beta$ -barrel fold of the back bone of interleukin 1 $\beta$ . The protein lacks disulfide bridges and consists of 12 antiparallel  $\beta$ -strands arranged into a  $\beta$ -barrel. There is a short  $3_{10}$  helix, which is unstable and exposed to the solvent.

nance spectroscopy, circular dichroism and fluorescence. The protein has been shown to refold kinetically through a significantly populated, compact intermediate state with structural properties resembling that of a molten globule state.<sup>269</sup> In the intermediate state the protein is found to regain 80% to 90% of the native secondary structure, but little or no stable tertiary structure. More than 90% of the ellipticity (corresponding to the secondary structure) is regained within the burst phase of folding (~10 ms). Analysis of the fluorescence data revealed the formation of three distinct intermediates during the refolding of the protein. A series of activation energy barriers appear to exist on the folding pathway of interleukin 1 $\beta$ . This feature probably leads to the accumulation of several intermediates along the refolding pathway. The fastest phase in the refolding of interleukin 1 $\beta$  corresponds to the formation of nearly complete native  $\beta$ -sheet structure. This event is paralleled by a major collapse of the polypeptide chain, as indicated by the change(s) in the environment of sole tryptophan residue and extensive hydrophobic cluster formation conducive for ANS binding. The native high-density packing of non-polar residues within the hydrophobic core forms on a time scale of about 20 mins.<sup>269</sup> Interestingly, quench-flow hydrogen-deuterium exchange experiments reveal that no stable backbone hydrogen bonds were formed during the dead time of the quench-flow apparatus. Complete backbone hydrogen bonding in the protein occurs only after 25 seconds of refolding. It is clear from the NMR data that the backbone interactions observed by far UV CD are not stable either due to local breathing or sliding of one strand relative to the other, resulting in the rapid rupture and formation of hydrogen bonds. Varley et al.<sup>269</sup> suggest that folding of this protein to its native structure involves the rapid formation of the  $\beta$ -sheet structure around a non-polar core which is consequently followed by much slower stabilization of the secondary structural elements accompanied by gradual final tight packing of the core groups external to the  $\beta$ -sheets. Based on the experimental results obtained in the kinetics of refolding of interleukin 1 $\beta$ , Gronenborn and Clore<sup>270</sup> rationalized the 'hydrophobic zipper' model proposed by Dill et al.<sup>271</sup> The 'hydrophobic zipper' model proposes that initiation of folding occurs by the grouping together of hydrophobic side chains which are closely positioned in the sequence. The grouping of the non-polar side chains is postulated to occur like the zipping of the zipper. In interleukin 1 $\beta$ , the initial zipper is proposed to be made up from the strands 6 and 7, with the other  $\beta$ -strands arranging around there.<sup>271</sup> Interestingly, the location and distribution of the non-polar side chains in interleukin 1 $\beta$  favors this proposal.<sup>270</sup>

The study of Heidary et al.<sup>272</sup> provides further insights into the folding mechanism of interleukin 1 $\beta$ . An attempt was made to characterize the nature of intermediates that accumu-

late during the refolding of interleukin 1 $\beta$ . A partially folded state was observed to be formed with a relaxation time of about 126 ms. of refolding at pH 5.0. Refolding of the protein monitored by stopped-flow CD at 230 nm reveals that all the expected amplitude is recovered in two observable phases.<sup>272</sup> The first phase accounts for 60% of the expected amplitude change and with a relaxation time of 152 ms. The slow phase has a relaxation time of about 43 s. The rate constants of both the detectable phases agree well with the data obtained from fluorescence studies.<sup>272</sup> Unlike the study of Varley et al.<sup>270</sup> wherein secondary and tertiary structural interactions were shown to develop on different scales, Heidary et al.<sup>272</sup> report that secondary and tertiary structural interactions are acquired simultaneously. The observed disparity between the two studies has been attributed to the difference(s) in pH of the refolding buffer used in these studies. The results obtained by Heidary et al.<sup>272</sup> clearly suggest that interleukin 1 $\beta$  refolds with a simple sequential model. Data obtained from pulse-labeling hydrogen exchange and electrospray ionization mass spectrometric analysis unambiguously suggest that native state of interleukin- $\beta$  proceeds through an obligatory intermediate.<sup>272-274</sup>

#### Apo-pseudoazurin

Apo-pseudoazurin is a 123 amino acid residue protein isolated from *Thiosphaera pantotropha*. The protein has complex double wound Greek key topology.<sup>275</sup> Interestingly, apo-pseudoazurin lacks disulfide bonds and tryptophan residues. The protein has a  $\beta$ -sandwich structure with 10  $\beta$ -strands (Fig. 16, Ref. 275). At the C-terminal end there are two short helices. Apo-pseudoazurin possesses eight proline residues with seven in the *trans* and one in the *cis* conformations. The

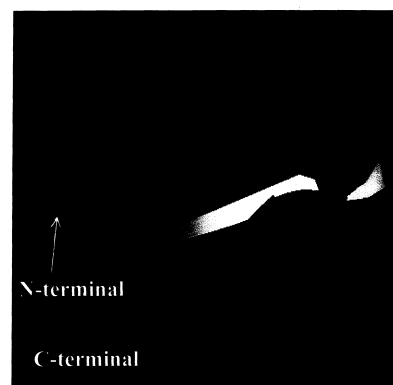


Fig. 16. Representation of the backbone folding of apo-pseudoazurin. The secondary structure of the protein includes 10  $\beta$ -strands arranged into a sandwich-type structure. There are two short helices at the C-terminal end of the molecule.

complex fold of the protein attracts the attention of protein folders since it seems unlikely that the non-local secondary structure (observed in the protein) could from before substantial native tertiary contacts are established. The complex motif is postulated to form *via* early for formation of an extended  $\beta$ -hairpin or hydrophobic zippering of amino acid side-chains. Capaldi et al.<sup>276</sup> recently investigated the folding events in the refolding of apo-pseudoazurin.

To understand folding of protein in the absence of proline isomerization, Capaldi et al.<sup>276,277</sup> employed a double jump technique wherein folding was initiated from a completely denatured state(s) in which all the eight proline residues are in their native isomeric state. The folding kinetics of apo-pseudoazurin as monitored by far UV CD spectroscopy (using the double jump approach) shows that about 50% of the CD signal of the native protein is regained within the dead time of the experiment, indicating rapid formation of an intermediate during folding. Generation of the native CD signal occurs in a single phase. In order to understand the conformation of the intermediate formed in the very early stages of folding, far UV CD employing the double jump technique was collected at various wave lengths. Comparison of this data with that of the protein in the native and unfolded states shows that the kinetic intermediate has substantial ellipticity in the far UV CD, suggestive of a mixture of  $\beta$ -sheet and random coil conformations.<sup>276</sup> The urea dependence of the unfolding and refolding kinetics of the protein was investigated to examine the accumulation of intermediate(s) along the folding/unfolding pathways of the protein. Strangely, the Chevron plot for apo-pseudoazurin showed no change of slope in the refolding limb. It has been demonstrated that transient aggregation can mimic the population of an intermediate in kinetic refolding experiment.<sup>276</sup> This phenomenon is observed at lower denaturant concentrations, and disaggregation of transient oligomers by denaturants makes the folding limb of the Chevron plot deviate from linearity. However, Capaldi et al.<sup>276</sup> opine that an intermediate populated in the dead time of apo-pseudoazurin could be unusually stable and denaturing close to the midpoint of the equilibrium unfolding transition. In such an event, the refolding limb of the Chevron plot is expected to be linear despite the accumulation of the intermediate in the refolding pathway. The high stability of the folding intermediate of pseudo-azurin strongly supports this proposal. Despite the lack of deviation from linearity of the Chevron plot, there is overwhelmingly strong experimental support for the occurrence of intermediate state(s) during the refolding of pseudo azurin. Capaldi et al.<sup>276</sup> examined whether the observed intermediate represents an on-pathway, productive intermediate. Filling the kinetic data to the on or off pathway model, it is found that apo-pseudoazurin refolds *via* an inter-

mediate which is obligate and on the 'folding pathway'.

### Cardiotoxin II

Cardiotoxin III (CTX III) is a small molecular weight (~7.0 kDa), all  $\beta$ -sheet protein isolated from the venom of the Taiwan cobra (*Naja naja atra*).<sup>278-283</sup> The protein presents a three-finger shaped appearance with three loops projecting from a globular head. The head portion of the molecule is cross-linked by four disulfide bonds.<sup>279,284-286</sup> The protein lacks tryptophan residue(s). The secondary structural elements in the protein include five  $\beta$ -strands arranged in an antiparallel fashion into double and triple stranded  $\beta$ -sheet domains (Fig. 17, Ref. 184). Secondary structure prediction analysis using various algorithms reveals that no segment of the amino acid sequences of CTX III has a propensity to adopt helical conformation.<sup>287-291</sup> Interestingly, addition of heliogenic solvents such as 2,2,2-trifluoroethanol (TFE) induces helical conformation in the protein.<sup>287,288</sup> These results are surprising considering the fact that the helix inducing tendencies of TFE have been believed to be specific.<sup>289</sup> TFE has been shown to induce helical conformation in only those portion(s) of the sequence(s) of proteins/peptides that are either helical or have a strong sequence propensity to adopt helical conformation.<sup>292</sup> Shiraki et al.<sup>292</sup> examines the TFE-induced conformational transitions on the native structure of more than twenty proteins. The results of this study showed that even predominantly  $\beta$ -sheet proteins such as  $\beta$ -lactoglobulin and concanavalin A show high helical propensity in TFE.<sup>292</sup> However, the induction of helix in these proteins is strongly correlated to the high propensity of the amino-acid sequence to adopt helical conformation. Jayaraman et al.<sup>288</sup> investigated the effects of TFE on CTX III in its native, denatured and

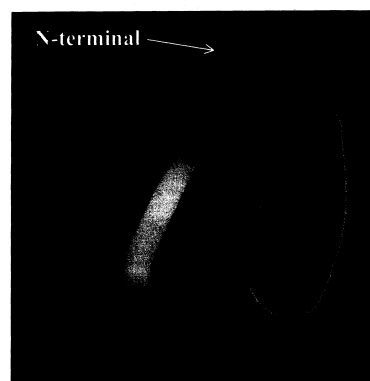


Fig. 17. MOLSCRIPT representation of the backbone folding of cardiotoxin III. It is a ~7 kDa, three-finger shaped protein. The protein has five  $\beta$ -strands arranged in an antiparallel fashion into double and triple stranded antiparallel  $\beta$ -sheet domains.

disulfide reduced denatured states. Results of this study revealed that the structural transitions induced by TFE in CTX III in its native state and denatured states are similar. About 25-30% helical conformation is induced in the protein in 90% (v/v). Surprisingly, helical conformation is induced in CTX III even in 20% (v/v) TFE upon reduction of the native disulfide bonds in the protein. These results suggest that helix-inducing effects of TFE could be non-specific. Sivaraman et al.<sup>289,293</sup> using S-carboxymethylated CTX III demonstrated that the native disulfide bonds in the protein effect the helix-induction by TFE. It is found that disruption of disulfide bridges in the protein results in the loss of tertiary structural interactions and consequently the helix-induction by TFE (in the protein) is more facile (Fig. 18). In another study, Sivaraman et al.<sup>289</sup> showed that the helix-induction by TFE is intricately linked to drastic destabilization of native tertiary structural interactions in CTX III. In addition, Kumar et al.<sup>294</sup> recently probed the TFE induced unfolding of CTX III under acidic conditions. The TFE-induced unfolding process is shown to be completely reversible. Using a variety of biophysical techniques, it was found that CTX III exists in a 'molten globule' like state in 80% (v/v) TFE under acidic conditions. Two-dimensional NMR studies revealed that the protein in its 'molten globule' like state has most of the native  $\beta$ -sheet secondary structural elements intact.<sup>294</sup> However, it is still not clear if this intermediate state is on/off the unfolding pathway of CTX III.

Partially structured states in CTX III have also been characterized in the equilibrium-unfolding pathway of CTX III under several conditions.<sup>295-298</sup> Jayaraman et al.<sup>298</sup> recently identified a partially structured intermediate which is populated during the thermal unfolding pathway of CTX III at pH

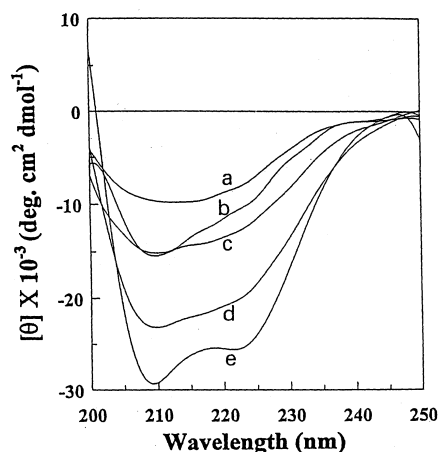


Fig. 18. Far UV CD curves of denatured and disulfide reduced CTX III at various concentrations of TFE. a - 0%; b - 30%; c - 40%; d - 70% and e - 90% (v/v).

4.0. Interestingly, the protein in its partially structured state is stable at high temperature ( $\sim 90^\circ\text{C}$ ). Similarly, Sivaraman et al.<sup>297</sup> characterized a stable, partially structured state in the 2,2,2-trichloroacetic acid (TCA)-induced unfolding pathway of CTX III. The partially structured intermediate (in 3% (w/v) TCA) exhibits strong affinity to the hydrophobic dye, ANS. Fourier transform IR, fluorescence and CD analysis revealed that CTX III in the 'molten globule'-like state has lost most of the native tertiary structural interactions.<sup>297</sup> Results of the one and two-dimensional NMR experiments indicate that about 65% of the native  $\beta$ -sheet structural contacts are maintained in the partially structured state of CTX III in 3% TCA (Fig. 19). Interestingly, most of the interactions stabilizing the triple stranded  $\beta$ -sheet domain are found to be unperturbed. However, the  $\beta$ -strands comprising the double strand are completely unfolded in the acid-induced partially structured state.<sup>297</sup>

The kinetics of refolding of CTX III from its guanidinium hydrochloride (GdnHCl) state was investigated recently by Sivaraman et al.<sup>299</sup> The unfolding  $\leftrightarrow$  refolding transitions induced by GdnHCl are reversible and are shown to follow two-state kinetics. Refolding kinetics monitored using

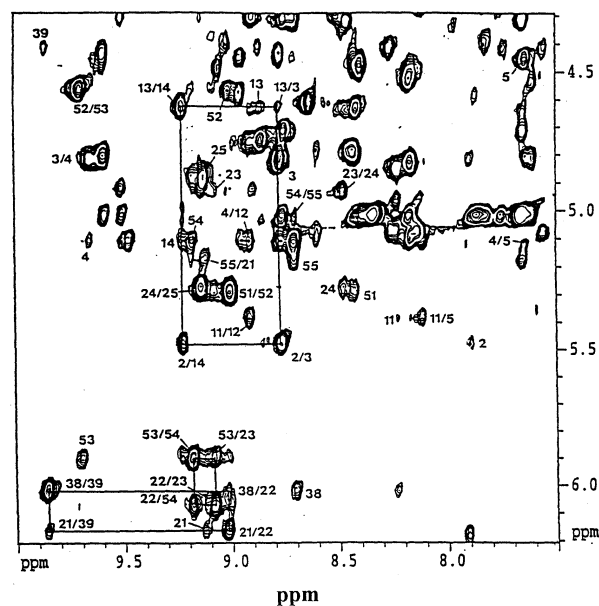


Fig. 19. NOESY spectrum (200 ms mixing period) of CTX III in 3% w/v TCA depicting NOEs in the fingerprint region (NH,  $\text{C}^\alpha\text{H}$ ). The sequential interstrand cross-peaks (NH,  $\text{C}^\alpha\text{H}$ ) are labeled and the intrasidial cross-peaks are indicated according to the position of the corresponding amino acid residue(s) in the protein sequence. Most of the interactions stabilizing the triple-stranded  $\beta$ -sheet domain are intact in the 'molten globule' like state in 3% (w/v) TCA.

ANS (as an extrinsic probe) reveals that the protein folds through a hydrophobic collapse in the burst phase (<5 ms) of refolding (Fig. 20). Interestingly, the refolding kinetics followed by quenched-flow deuterium-hydrogen exchange of the amide protons reveals that the residues that possess refolding time constants less than 20 ms (fast folding residues) are polar residues concentrated in the triple-stranded  $\beta$ -sheet segment and the C-terminal loop.<sup>299</sup> The kinetics of refolding of CTX III monitored by far and UV CD reveals an overshoot in the 214 nm and the 222 nm signal. Refolding kinetics of CTX III in the presence of TFE did not show further increase in the intensity of the 222 nm signal implying that the overshoot is not related to formation of non-native helix conformation during the refolding of the protein. It is suggested that the far UV 'overshoot' observed in CTX III originate due to the regain of asymmetry of the disulfide bridge(s) during the refolding of the protein.<sup>299</sup>

Quenched-flow hydrogen-deuterium studies on the refolding of CTX III showed that the protein folds within a time period of 125 ms. The average refolding time constant of the five  $\beta$ -strands comprising the double and triple stranded  $\beta$ -sheet domains in CTX III are: strand I = 34.8 ms, strand II = 35.1 ms, strand III = 23.6 ms, strand IV = 18.3 ms and strand V = 17.2 ms. Formation of strand II and strand I constituting the double stranded  $\beta$ -sheet segment appears to occur almost simultaneously. The average time constant of residues in the double stranded  $\beta$ -sheet is 35.0 ms.<sup>299</sup> The average time constant of residues comprising the triple stranded  $\beta$ -sheet segment is 19.7 ms. Among the three  $\beta$ -strands (strands III, IV and V), residues in strand V fold rapidly with an average time constant of 17.2 ms. In summary, the triple stranded  $\beta$ -sheet

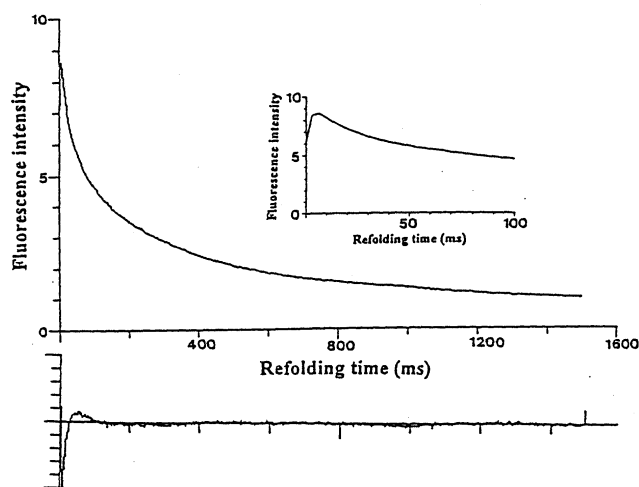


Fig. 20. Stopped-flow kinetics of refolding of CTX III monitored by ANS fluorescence. The inset clearly depicts the hydrophobic collapse that occurs in the burst phase of folding.

domain in CTX III folds faster than the double stranded  $\beta$ -sheet segment (Fig. 21). Recently, Sivaraman et al.<sup>286</sup> investigated the native state amide proton exchange kinetics of CTX III to understand if the average protection factor(s) and the time constant of refolding of various amide protons in CTX III are correlated. It is found that the most slowly exchanging portion constitutes the folding core of the protein.

### Cobrotoxin

Cobrotoxin (CBTX) isolated from the venom of the Taiwan cobra (*Naja naja atra*) is a small molecular weight all  $\beta$ -protein cross-linked by four disulfide bonds.<sup>300,301</sup> Cobrotoxin shares more than 45% sequence homology with CTX III. Comparison of the three-dimensional structures of CBTX and CTX III reveals that both the proteins are 'three-finger' shaped with three loops projecting from a globular head (Fig. 22, and Refs. 284, 300). CBTX, like CTX III, contains five  $\beta$ -strands arranged in an antiparallel fashion into double and triple stranded  $\beta$ -sheets.<sup>300</sup> The free energy of unfolding of CBTX is about  $2.25 \pm 0.5$  kcal/mole which is about half the values estimated for CTX III (4.21 kcal/mol). Sivaraman et al.<sup>301</sup> recently investigated the folding kinetics of CBTX using a variety of biophysical techniques. The secondary structure formation and hydrophobic collapse occur as distinct events during the refolding of the protein. More than 80% of the total expected amplitude change is realized within about 30 ms after refolding is initiated.<sup>301</sup> Quenched-flow deuterium-hydrogen exchange data obtained at various pulse labeling times reveals that changes in the amide proton(s) protection are mostly complete within 75 ms of refolding (Fig. 23). The formation of tertiary structural interactions during the refolding of CBTX was monitored by two different spec-

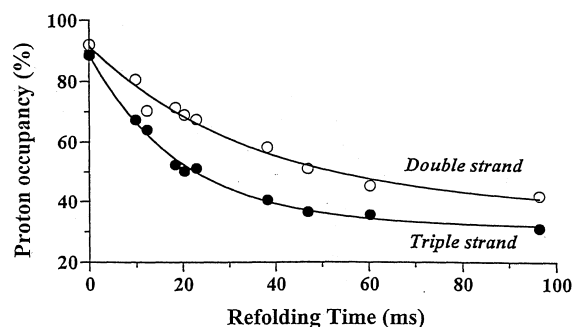


Fig. 21. Time course for the protection of amide protons from exchange of the residues involved in double and triple stranded  $\beta$ -sheets during the refolding of CTX III. Non-linear least square fits of the data yield values of 19.7 ms (filled circle) and 35.0 ms (open circles) for the average time constants of folding of the triple and double stranded  $\beta$ -sheet domains, respectively.

troscopic probes. The 284 nm ellipticity practically remains unchanged even up to 100 ms, suggesting that little or no tertiary structural contacts develop in the protein during the time frame wherein complete secondary structural interactions are formed. In addition, less than 5% of the final amplitude of the ANS fluorescence is regained in the first 75 ms of refolding of CBTX, indicating that no or very meager extent of tertiary structure develops within this time scale.<sup>301</sup> The ANS fluorescence shows a steep increase to a maximum value at 200 ms (Fig. 24). It is rather surprising to find that folding of a small protein such as cobrotoxin proceeds through the occurrence of

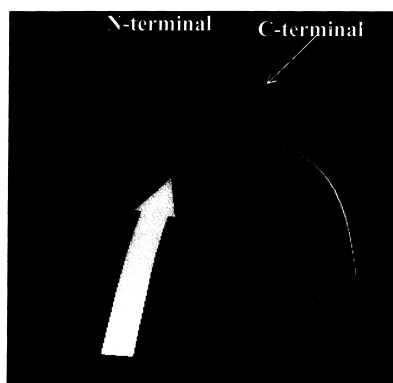


Fig. 22. Back bone folding of cobrotoxin. The anti parallel, double and triple stranded  $\beta$ -sheet domains in the protein are tethered by the presence of four disulfide bonds (not shown). The protein has a single tryptophan residue, which could be successfully exploited to monitor the folding of the protein.

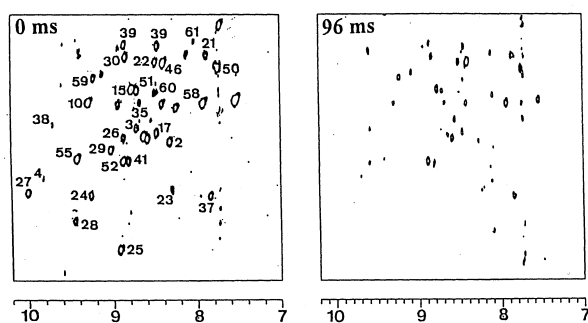


Fig. 23. Magnitude COSY spectra of CBTX samples prepared by quenched-flow hydrolysis exchange methods, at two different time periods. The NH- $^1$ H cross-peaks observed in the spectra after 96 ms of refolding represent residues in the unstructured portions of the CBTX molecule. The labeled cross-peaks (at 0 ms of refolding) represent the residues in the secondary structure.

distinct structural events. The refolding pathway of CBTX is an unique example wherein secondary structure formation and hydrophobic collapse occur as well segregated events. The events in the refolding kinetics of CBTX would help shed the general notion that folding of small proteins do not involve the occurrence of distinct intermediate(s).<sup>301</sup>

### Fibroblast Growth Factor

Fibroblast growth factor (FGF) is a ~17 kDa, heparin binding, all  $\beta$ -sheet protein.<sup>302</sup> The protein has free cysteine residues but lacks disulfide bonds. Interestingly, the protein has a single tryptophan residue, which could serve as a useful probe to monitor the folding/unfolding of the protein. The protein folds according to a  $\beta$ -trefoil motif.<sup>303</sup> Human acidic FGF (hFGF), which is the most well studied FGF analog, has six  $\beta$ -strand pairs, five of them with hair pin structure and another that, topologically equivalent to the other five and sometimes referred to as the sixth hair pin, is not as such.<sup>303,304</sup> Three of these pairs form a six-stranded barrel structure and the other three are in a triangular array that caps the barrel (Ref. 304, Fig. 25). In general, the structure of hFGF bears a gross similarity with that of interleukin 1 $\beta$ .<sup>305,306</sup>

As mentioned earlier, FGF is a heparin binding protein and is known to possess a distinct anion binding site. Burke et al.<sup>306</sup> showed that several polyanions could bind to hFGF and stabilize its structure against urea and heat denaturation. Sanz and Gallego recently investigated the acid-induced unfolding of hFGF.<sup>307</sup> They demonstrate that hFGF exists in a partially structured state at pH 4.0. The protein in its partially structured state is shown to interact weakly with the hydrophobic probe N-phenyl-1-naphthylamine indicating a relatively high level of structure which did not fit into the classical molten-globule category.<sup>307</sup> Interestingly, this intermediate is

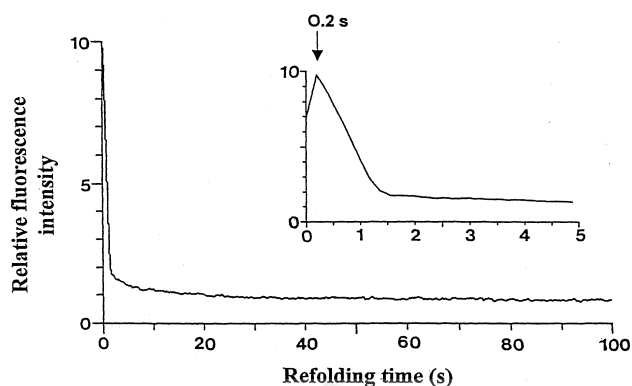


Fig. 24. Changes in the ANS fluorescence during the refolding of CBTX. The inset figure represents the changes in the ANS fluorescence in the first 1 s of refolding of the protein.

shown to interact with liposomes suggesting that it might represent a membrane translocation-competent form. Similar conclusions were made by Mach and Middaugh<sup>308</sup> based on their membrane interaction studies with hFGF. hFGF is shown to interact with negatively charged but not neutral phospholipid unilamellar vesicles at acidic pH, inducing bilayer disruption. The negatively charged lipid bilayers appear to interact with partially structured states of hFGF by preferential binding of both its apolar and charged surfaces to complementary regions of the lipid bilayer.

Dabora et al.<sup>309</sup> investigated the effects of polyanions on the refolding of hFGF. In addition, the effect(s) of temperature on the kinetics of refolding of hFGF were also examined in the presence and absence of several polyanions. The refolding kinetics monitored by the change in tryptophan fluorescence reveals that complete folding of protein (hFGF) takes more than 50s. The rate of folding was surprisingly dependent on protein concentration, with the rate doubling as the concentration dependence was raised from the 20/ $\mu\text{g}/\text{ml}$  at 25 °C. However, in the presence of EDTA this concentration dependence of refolding is abolished. This aspect probably suggests a complex kinetic pathway for hFGF in the presence of heavy metal ions, which could potentially oxidize the free thiol available in the protein. The rate of refolding of hFGF is shown to be significantly increased in the presence of various polyanions. Interaction of polyanions with hFGF is shown to decrease the activation energy for folding. Additionally, the presence of molecular chaperones and peptidyl prolyl isomerase does not seem to affect the refolding rates of hFGF from its unfolded state(s).

### $\beta$ -lactoglobulin

$\beta$ -lactoglobulin is isolated from milk sources and is a 18

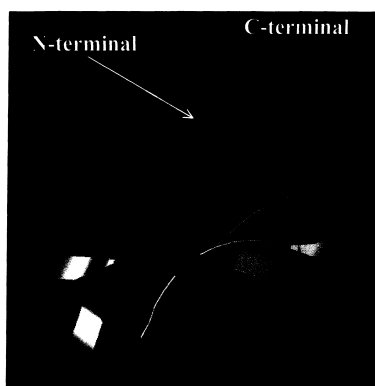


Fig. 25. The structure of the human acidic fibroblast growth factor. The protein has a single tryptophan residue and lacks disulfide bonds. The protein has 12  $\beta$ -strands arranged into a  $\beta$ -barrel structure.

kDa globular protein with two disulfide bridges.<sup>310</sup> The protein is generally a monomer but it dimerizes in alkaline conditions due to the inter-molecular disulfide bond formation between the free thiol groups located in each monomer. X-ray structure of  $\beta$ -lactoglobulin indicates that the protein is a  $\beta$ -barrel composed of nine  $\beta$ -strands and one small  $\alpha$ -helical segment (Fig. 26). Interestingly, comparison of the x-ray crystallographic structures of plasma retinol binding protein and  $\beta$ -lactoglobulin bear remarkable similarity and are categorized in the lipocalin family.<sup>311,312</sup> Bovine  $\beta$ -lactoglobulin represents an interesting example of context-dependent secondary structure induction.<sup>313</sup> Secondary structure prediction analysis reveals a high  $\alpha$ -helical preference, which is retained for short fragments (Shiraki et al.).<sup>161</sup>

Molinary et al.<sup>314</sup> recently studied the pH dependent conformational status of bovine  $\beta$ -lactoglobulin. It is found that the protein exists in a partially structured state at pH 2.0. The protein is found to exist in a monomeric state under these conditions.<sup>314</sup> NMR characterization of the pH induced partially structured state of  $\beta$ -lactoglobulin provides evidence of extensive flexibility of the aromatic residues. In general, the protein contains a sub-domain with a highly ordered  $\beta$ -sheet core, while the rest of the molecule exhibits an increased flexibility, giving rise to locally ordered structures.<sup>314</sup> NMR studies of the intermediate population in the acid-induced unfolding pathway revealed a conserved hydrophobic cluster and is believed to play an important role in the stability of the protein. This hydrophobic cluster is shown to have native-like structure. In a related study, Fujiwara et al.<sup>315</sup> reported the accumulation of an equilibrium unfolding of equine  $\beta$ -lactoglobulin using guanidinium hydrochloride. The interme-

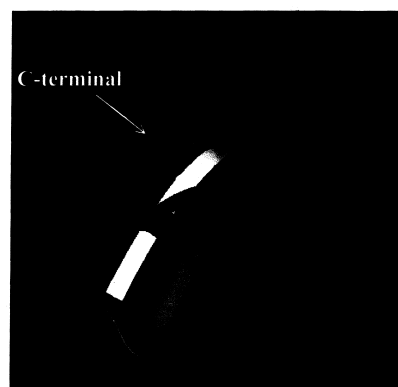


Fig. 26. Depiction of the backbone folding of  $\beta$ -lactoglobulin. The  $\beta$ -barrel structure of the protein is constituted by nine  $\beta$ -strands arranged in an antiparallel fashion. A small helical segment spanning six residues could also be observed. Secondary structure prediction analysis of the protein reveals a high  $\alpha$ -helical preference.

intermediate state has structural features akin to the 'molten globule' state. Interestingly, stable non-native helical segments were induced in the accumulated intermediate.<sup>315</sup> In contrast, urea-folding studies of  $\beta$ -lactoglobulin is shown to be an apparent two-state transition and no non-native  $\alpha$ -helical intermediate(s) were observed.<sup>315</sup> Dufour et al.<sup>316,317</sup> investigated the pressure-induced unfolding of  $\beta$ -lactoglobulin at two pHs. At neutral pH, the protein transitions under increasing pressure involves three states, native protein (at low pressure), denatured protein (under high pressure) and denatured state even after pressure release. Pressure-induced denaturation data at low pH (pH ~ 2) shows a marked increase in protein stability suggesting that the protein, while preserving the same secondary structure, can adopt various tertiary structures displaying different susceptibility to denaturing agents at different pH values. This assumption is supported by the differential pH dependent susceptibility of human lactoglobulin to pepsin hydrolysis.<sup>316,317</sup>

Kuwajima et al.<sup>318</sup> studying the refolding  $\beta$ -lactoglobulin reported the accumulation of a transient intermediate in the burst phase of refolding from its GdnHCl unfolded state.<sup>318</sup> Substantial secondary structure but no or little tertiary structural interactions were found to be present in the transient intermediate. Recently, Kuwajima and co-workers reinvestigated the folding kinetics of bovine  $\beta$ -lactoglobulin using circular dichroism and absorption measurements.<sup>319</sup> The burst phase in intermediate showed more intense ellipticity signals in the far UV region than the native state. The observed increase in ellipticity is attributed to the formation of non-native helical structure. Interestingly, the burst phase in intermediate is shown to share many common structural features with 'molten globule' state realized in the equilibrium intermediates, which include (1) presence of secondary structure and lack of tertiary contacts, and (2) formation of hydrophobic core. Residual  $\beta$ -structure were found in the disulfide cleaved, 4 M GdnHCl – denatured unfolded state(s) sample.<sup>319</sup> The persistent  $\beta$ -structure is a part of the native  $\beta$ -structure; it is believed that it ( $\beta$ -structure) could constitute the folding initiation site.<sup>319</sup> In a related study, Hamada et al.<sup>320-322</sup> studying the refolding kinetics of bovine  $\beta$ -lactoglobulin also reached a similar conclusion(s). They characterize the conformation of the burst phase-intermediate and suggest that the protein follow a 'non-hierarchical' model of folding and the non-native  $\alpha$ -helical structures are contemplated to play important roles in the folding of  $\beta$ -lactoglobulin. The burst phase in intermediate was characterized by the measurement of the wavelength dependence of the refolding kinetics.<sup>322</sup> The overshoot observed in the ellipticity values (in the far UV region) were shown not to be limited to ~220 nm but was also found across a broad range of the far UV region. To eliminate the possible artifacts

arising out of absorption of chromophores (in the far UV region), the refolding kinetics of the protein was examined in the presence of TFE. It was found that TFE stabilizes the helical segments formed in the burst phase in intermediate. Arai et al.<sup>323</sup> characterized the structural features of  $\beta$ -lactoglobulin using a variety of biophysical techniques including synchrotron x-ray scattering. These results reveal that  $\beta$ -lactoglobulin forms a compact globular structure within 30 ms of refolding. The radius of gyration within 100 ms of refolding was 1:1 times larger than that in the native state.<sup>323</sup> Their study indicated that both local and non-local interactions are dominant forces in the early stages of protein folding. Refolding of equine  $\beta$ -lactoglobulin from its urea-denatured state forms a 'molten globule' like intermediate in the burst phase of refolding.<sup>315</sup> The thermodynamic stability of the burst phase in intermediate estimated from the urea-concentration dependence of the burst phase amplitude is in good agreement with that of the equilibrium intermediate. Subramaniam et al.<sup>324</sup> reported  $\beta$ -lactoglobulin from its GdnHCl denatured state and monitored the regainment of the retinol binding and functional activity upon refolding of the protein. Interestingly, 10 times dilution of the denaturant does not allow the protein to regain the native conformation but completely restores its biological activity. These results raise an intriguing possibility that a fully folded structure (i.e. the global energy-minimum species) is not required for biological function of a protein. Haltori et al.<sup>325</sup> monitored the events in the refolding of  $\beta$ -lactoglobulin with monoclonal antibodies as probes. Complete renaturation was never realized under several renaturing conditions such as quicker or slower removal of the denaturant. These results are suggestive of some specific moiety(ies) in the protein molecule unable to return to the native conformation.<sup>325</sup>

### SH3 Domain Proteins

SH3 domains are integral parts of many signal transduction and cytoskeletal proteins and are believed to mediate a vast array of protein-protein interactions.<sup>326</sup> The abundance of numerous crystal and solution structures and their simplicity in terms of small size and lack of disulfide bonds, and bound co-factors render members of this class of proteins to be attractive models for understanding the rules of protein folding.<sup>327-329</sup> With the exception for variations in the loop regions, the SH3 domain proteins have a distinctive fold of two 3-stranded  $\beta$ -sheets packed orthogonally against each other to form a single hydrophobic core (Fig. 27). Viguera et al.<sup>330</sup> established the thermodynamic and kinetic analysis of the SH3 domain of spectrin. In principle, the folding and unfolding reactions of this SH3 domain were found to be described by the two-state model. No intermediate was observed to accumulate

by equilibrium denaturation monitored by fluorescence or circular dichroism methods.<sup>330</sup> A concomitant recovery of secondary and tertiary structure is observed during refolding. Farrow et al.<sup>331</sup> compared the backbone dynamics in equilibrium (in aqueous buffer) using multi-dimensional NMR methods. Analysis of the folding transition of the SH3 domain, which exists in a folded-unfolded equilibrium in non-denaturing conditions, indicated variability in the equilibrium constant for different residues in the molecule.<sup>331</sup> Protein in the folded form is shown to have as a compact structure, with uniform dynamic behavior. Comparison of the magnitudes of the values of the spectral density functions as well as coupling constant and NOE data reveals that the unfolded state(s) exists as an ensemble of relatively weak compact structures, with motional properties similar to the unordered regions in folded proteins. Zhang and Kay<sup>332</sup> further attempted to characterize the folded and unfolded states of an isolated N-terminal *src* homology 3 (domain) of *Drasphila drk*. NMR spectra of the proteins recorded in the folded and unfolded states simultaneously at neutral pH showed an approximate 1:1 ratio of protein conformations.  $\beta$ -turns like conformation were evident among the unfolded molecules of the protein.<sup>332</sup> A stretch of sequential amide-amide nuclear Overhauser effect cross-peaks could be located in a region corresponding to the  $\beta$ -strands in the native state. It could be worthwhile to debate whether the turn-like conformation identified in the unfolded states can be viewed as intermediate in the folding pathway. It is interesting to note that the SH3 domain of GRB2 is also known to exist in the equilibrium between folded and unfolded states at pH 3.5. NMR studies of the N-terminal of the GRB2 SH3 domain report extremely few or

no amide protons which exchange slowly with water, indicative of instability of the domain.<sup>332</sup> Zhang and Kay studied the effect of binding of a proline-rich peptide on the equilibrium between the folded and unfolded states.<sup>333</sup> It is found that the polyproline peptide stabilizes the native state. At an equimolar ratio of the SH3 domain to the peptide, 85% of the molecules exist in the folded state.<sup>333</sup> The results of the study bear a great deal of physiological significance. Given the large number of proteins which function by binding to other proteins, nucleic acids, or ligands, it is possible that a significant number of proteins *in vivo* require binding target ligands to form ordered and folded structures. It is interesting to note that sizeable populations of unfolded states of proteins are present *in vivo* under native conditions. In a recent report Blanco et al.<sup>334</sup> studied the NOEs for the unfolded states of wild type and mutants designed with a sequence with the highest predicted content of helical structure. The mutations were designed using the AGADIR algorithm and were tested in short peptides with modified sequences. Analysis of different mutant proteins did not show any clear relationship between peptide helical content and destabilization of the folded state. However, the kinetics of unfolding of the mutant proteins suggests that the unfolded states of the mutants are more compact than the wild type. These results demonstrate that existence of a high population of non-native secondary structural elements in the unfolded state(s) is compatible with the formation of the native structure. In an elegant study, Mok et al.<sup>335</sup> using NOEs obtained from multi-dimensional NMR techniques observed several long-range amide NOEs in the unfolded state(s) of the SH3 domain of *Drosophila drk*. These long-range amide NOEs were found to disappear upon addition of chemical denaturants, implying that substantial differences between the unfolded state(s) induced under physiological conditions and in the presence of chemical denaturants.<sup>335</sup>

Recently, Viguera et al.<sup>336</sup> analyzed the stability and folding kinetics of circularly permuted forms of the  $\alpha$ -spectrin SH3 domain. All possible circularly mutated proteins that differentiate two consecutive elements of secondary structure interact in the three-dimensional structure of the protein. In essence, each of the mutants has two separated sequences, which in the early stages of the folding process could be possibly involved in local interactions.<sup>336</sup> Exhaustive analysis of the possible permutations between secondary structural elements revealed that circularly mutated protein(s) folds to a conformation similar to the wild type. This study provides strong evidence to suggest that local interactions between existing secondary structural elements (for example,  $\beta$ -hairpins) are not guiding forces in the protein folding process.<sup>336</sup> In addition, it has been shown that thermodynamic stabilities of all the circular mutants are less than the wild type protein. Elaboration

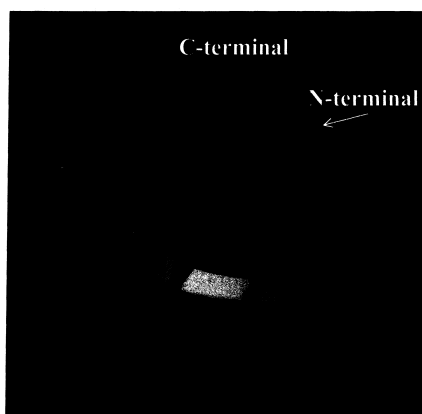


Fig. 27. Representative structure of the SH3 domain of  $\alpha$ -spectrin. The domain lacks disulfide bonds and has a distinctive fold of 2 three stranded  $\beta$ -sheets arranged orthogonally against each other to constitute a hydrophobic core.

rate protein engineering experiments reveal that the folding pathway of the protein has been changed in the circularly permuted proteins. It appears that although the order of secondary structure and the preservation of  $\beta$ -hairpin has no marked effect on the protein to refold to its native state, they critically influence the kinetics of the folding and unfolding pathways. Serrano and co-workers recently compared the energetics and structures of native and transition states of the wild types and circularly permuted mutants of the  $\alpha$ -spectrin SH3 domain.<sup>337</sup> Their study provides compelling evidence to suggest that proteins with similar amino acid composition and folds can have different transition states. It appears that the folding nucleus is not specific and presence of certain residues in the folding cluster are simply more probable but not essential. Even if one subset of folding routes dominates in a given sequence, an entirely different subset of routes may dictate folding in another.<sup>337</sup>

Viguera and Serrano investigated the influence of loop length and intramolecular diffusion on the protein folding process.<sup>338</sup> Circularly permuted mutants of  $\alpha$ -spectrin SH3 domain with different poly-glycine loop insertions (up to 10 glycine residues) showed that the glycine insertion does not affect the structure of the folded state. In addition, the structural characteristics of the denatured ensemble also did not change significantly by the insertion of the glycine loop. However, the apparent energy barrier between the folded and unfolded states increased linearly with the length of the glycine loop inserted.<sup>338</sup> Insertion of the glycine loop produced no marked conformational changes. The free energy change for a 10-glycine insertion is rather small ( $\sim 0.8$  kcal mole<sup>-1</sup>). Interestingly, loop elongation is found to accelerate unfolding but slows down the folding rate. Although the effects on both unfolding and refolding rate constants are similar in magnitude, the dependence of their natural logarithm on loop length is different. In the unfolding phase, the effect is non-linear and tends to saturate. In contrast, in the folding phase, the loop length and rate constant almost increase linearly. The results of this study suggest that the distance among interacting residues is important in overcoming the high energy transition state barrier and thus confers a role for intermolecular diffusion in the protein folding process.

Prieto et al.<sup>339</sup> addressed the question of the relative importance of secondary versus tertiary structural interactions for stabilizing and the folding process. Mutations were designed which introduce a strong non-native helical propensity in the N-terminus of the  $\alpha$ -spectrin SH3 domain.<sup>339</sup> Although the mutant proteins had similar three-dimensional structures as the wild type, they were less stable and had less cooperative folding transitions. A very good correlation could be found from the refolding ' $m_f$ ' values and the helical ten-

dency. As the mutants exhibit a two-state transition with no folding intermediate(s), the decrease observed in the ' $m_f$ ' value can not be completely explained on the basis of a decrease in the solvent accessibility of the denatured state.<sup>339</sup> The stabilization of the non-native helical structures in the SH3 domain probably leads to compaction of the denatured state.

The folding kinetics of the SH3 domain of P13 kinase was investigated recently using a variety of biophysical techniques.<sup>340</sup> In contrast to the  $\alpha$ -spectrin SH3 domain, the folding kinetics of the P13 kinase SH3 domain is biphasic. The slow phase corresponds to the incorrectly configured *cis* proline residue(s). Real-time NMR analysis of the folding kinetics of the protein shows that the species populated in the initiation stages of folding exhibit poor chemical shift dispersion similar to the unfolded state in 6 M GdnHCl. PI3-SH3 domain folds slowly with a time constant of 2.8 seconds at 20 °C.<sup>340</sup> It is interesting to note that with the exception of the P13-SH3 domain, all other SH3 domains studied so far fold extremely rapidly with the time constant values ranging from 1 to 340 ms. These results suggest that there is no strict relationship between the absence of folding intermediates and rapid folding. In addition, slow folding may not be a consequence of occurrence of kinetic traps.<sup>340</sup>

Experimental studies on the folding aspects of *src* SH3 domain were recently complemented by molecular dynamics simulations.<sup>341</sup> About 30 independent, high temperature molecular dynamics simulations of the *src* SH3 domain unfolding were monitored. The cumulative analysis of the observed trajectories revealed a hierarchy of events in the unfolding process. The contacts between the N- and C-terminal strands are the first to disappear.<sup>341</sup> The hydrogen-bonding network involving the distal  $\beta$ -hairpin and the diverging turn persists even after most of the native structural contacts are lost. Although the overall hierarchy of structure loss is similar in the simulations and in experiment, differences are more dramatic in the experimental results. The experimental  $\phi$  values are close to zero for the first and last  $\beta$ -strand, and near one in the distal loop beta hairpin. The computed  $\phi$  values for the same structured regions (in the protein) from simulations are on average 0.4 and 0.8, respectively.<sup>341</sup>

### Common Features in the Folding of $\beta$ -Sheet Proteins

A survey of the folding kinetics of  $\beta$ -sheet proteins studied thus far reveals a broad range of rate constants. Large multi-domain  $\beta$ -sheet proteins such as interleukin 1 $\beta$  appear to fold relatively slowly. In contrast, small single domain proteins such as cardiotoxin III, cobrotoxin and SH3 domain protein fold extremely rapidly on time scales less than 500 ms. Cold shock protein B is an extreme example of a rapidly fold-

ing  $\beta$ -sheet protein ( $k = 1070 \text{ s}^{-1}$ ). The single domain  $\alpha$ -helix and  $\beta$ -sheet proteins are observed to fold on similar time scales. In general it appears that  $\alpha$  and  $\alpha+\beta$  proteins and all  $\beta$ -sheet proteins use different principles to reach the activated state in their folding reactions. The dichotomy could be best understood by noting that helices are stabilized primarily by local interactions and hence conform very easily and rapidly. In most of the helical proteins, folding is initiated at helices.  $\beta$ -sheets on the other hand are stabilized by non-local interactions and possibly almost the entire  $\beta$ -sheet is necessary to reach the required activated state. There are only a few productive folding routes since the  $\beta$ -sheet structure is more difficult to form and stabilize. For this reason folding pathways are well conserved among related  $\beta$ -sheet proteins. In addition, the  $\beta$ -sheet proteins have well a defined activated state which closely resembles the native state. Interestingly, the folding pathways are quite stringent and cooperative and can not be easily altered by gross mutations. In addition, the fact that  $\beta$ -sheet proteins are not at a disadvantage of attaining their native state(s) suggests that early restriction of conformational freedom does occur in  $\beta$ -sheet. A better understanding of the very early stages of folding of  $\beta$ -sheet proteins in the future would help immensely in the rational design of  $\beta$ -sheet proteins with desired function(s).

#### ACKNOWLEDGEMENTS

We would like to thank all our collaborators and coworkers who have significantly contributed to research on protein folding. We express our appreciation to the Regional Instrumentation center at Hsinchu (supported by the National Science Council, Taiwan) for the 600 MHz NMR facility. We would also like to thank the National Science Council, Taiwan for funding various projects (NSC 88-CPC-B007-001, NSC 89-2311-B007-020 and NSC 89-2313-M007-012).

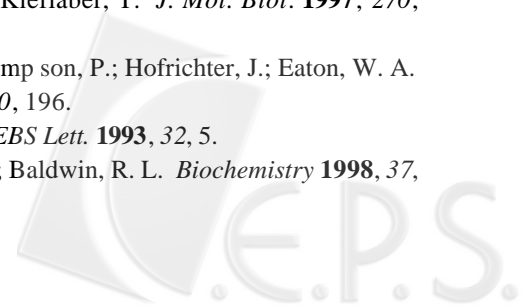
Received December 21, 1999.

#### Key Words

Levinthal Paradox; Molten globule; Intermediate;  $\beta$ -sheet protein.

#### REFERENCES

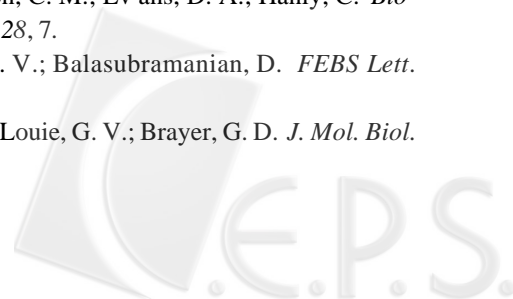
- Anfinsen, C. B.; Haber, E.; Sela, M.; White, F. H. Jr. *Proc. Natl. Acad. Sci., USA* **1961**, *47*, 1309.
- Anfinsen, C. B. *Science* **1973**, *181*, 223.
- Dobson, C. M.; Sali, A.; Karplus, M. *Angew. Chem.* **1998**, 868-893.
- Dalby, P. A.; Oliverberg, M.; Fersht, A. R. *J. Mol. Biol.* **1998**, *276*, 625.
- Hill, T. L. *Proc. Natl. Acad. Sci., USA* **1976**, *73*, 679.
- Fersht, A. R.; Itzhaki, L. S.; ElMarsy, N. F.; Matthews, J. M.; Otzen, D. E. *Proc. Natl. Acad. Sci., USA* **1994**, *91*, 10426.
- Chamberlain, A. K.; Marqusee, S. *Structure* **1999**, *5*, 859.
- Pande, V. S.; Grosberg, A. Y.; Tanaka, T.; Rokhsar, D. S. *Curr. Opin. Str. Biol.* **1998**, *8*, 68.
- Lazaridis, T.; Karplus, M. *Science* **1997**, *278*, 1928.
- Cook, K. H.; Schmid, F. X.; Baldwin, R. L. *Proc. Natl. Acad. Sci., USA* **1979**, *76*, 6157.
- Colton, W.; Roder, H. *Nat. Struc. Biol.* **1996**, *3*, 1019.
- Milla, M. E.; Brown, B. M.; Waldburger, C. D.; Saur, R. T. *Biochemistry* **1995**, *34*, 13914.
- Creighton, T. E. *Curr. Biol.* **1997**, *7*, R380.
- Baldwin, R. L. *Curr. Opin. Struc. Biol.* **1993**, *3*, 84.
- Elezer, D.; Yao, J.; Dyson, H. J.; Wright, P. E. *Nat. Struc. Biol.* **1998**, *5*, 148.
- Bryngelson, J. D.; Onuchich, J. N.; Socci, N. D.; Wolynes, P. G. *Protein Struc. Funct. Genet.* **1995**, *21*, 167.
- Dill, K. A.; Chan, H. S. *Nat. Struc. Biol.* **1997**, *4*, 10.
- Marmorino, J. L.; Lethi, M.; Pielak, G. J. *J. Mol. Biol.* **1998**, *275*, 379.
- Wu, L.; Kim, P. S. *J. Mol. Biol.* **1998**, *280*, 175.
- Carlisch, A.; Karplus, M. *Proc. Natl. Acad. Sci., USA* **1994**, *91*, 1746.
- Li, A.; Dagget, V. *Proc. Natl. Acad. Sci., USA* **1994**, *91*, 10430.
- Klimov, D. K.; Thirumalai, D. *Phys. Rev. Lett.* **1997**, *79*, 317.
- Matthews, C. R. *Methods Enzymol.* **1987**, *154*, 498.
- Yeh, S. R.; Rosseau, D. L. *Nat. Struc. Biol.* **1998**, *5*, 222.
- Ikeguchi, M.; Kuwajima, K.; Mitani, M.; Sugai, S. *Biochemistry* **1986**, *25*, 6965.
- Kay, M. S.; Baldwin, R. L. *Nat. Struc. Biol.* **1996**, *3*, 439.
- Mitraki, A.; Danner, M.; King, J.; Seckler, R. *J. Biol. Chem.* **1993**, *268*, 20071.
- Schmid, F. X.; Blaschek, H. *Eur. J. Biochem.* **1981**, *114*, 111.
- Zaidi, F. N.; Nath, V.; Udgaonkar, J. B. *Nat. Struc. Biol.* **1997**, *4*, 1016.
- Wildegger, G.; Kieflaber, T. *J. Mol. Biol.* **1997**, *270*, 294.
- Munov, V.; Thompson, P.; Hofrichter, J.; Eaton, W. A. *Nature* **1997**, *390*, 196.
- Fersht, A. R. *FEBS Lett.* **1993**, *32*, 5.
- Goldberg, J. M.; Baldwin, R. L. *Biochemistry* **1998**, *37*,



- 2556.
34. Matthews, C. R. *Ann. Rev. Biochem.* **1993**, 62, 653.
  35. Schulman, B. A.; Kim, P. S.; Pobson, C. M.; Redfield, C. *Nat. Struct. Biol.* **1992**, 4, 630.
  36. Peng, Z.; Kim, P. S. *Biochemistry* **1994**, 33, 2136
  37. Doyle, R.; Simons, K.; Qian, H.; Baker, D. *Protein Struct. Funct. Genet.* **1997**, 29, 282.
  38. Levinthal, C. in *Mossbauer Spectroscopy in Biological Systems*; Debrunner, P.; Tsibris, J. C. M.; Miinck, E., Ed.; University Press: Urbana, **1969**; pp 22-32.
  39. Shortle, D.; Chan, H. S.; Dill, K. A. *Protein Sci.* **1992**, 1, 200.
  40. Shrivastava, J.; Visheshwara, S.; Cieplak, M.; Maritan, A.; Banavar, J. R. *Proc. Natl. Acad. Sci., USA* **1995**, 92, 9206.
  41. Levinthal, C. *J. Chem. Phys.* **1968**, 65, 44.
  42. Oliverberg, M.; Tan, Y. J.; Silow, M.; Fersht, A. R. *J. Mol. Biol.* **1998**, 277, 933.
  43. Itzahaki, L. S.; Otzen, D. E.; Fersht, A. R. *J. Mol. Biol.* **1995**, 254, 260. *Nat. Struct. Biol.* **1998**, 5, 714.
  44. Du, R.; Pande, V. S.; Grosberg, A.; Yu, T.; Tanaka, T.; Shakhovich, E. *J. Chem. Phys.* **1998**, 108, 334.
  45. Fersht, A. R. *Curr. Opin. Struct. Biol.* **1997**, 7, 3.
  46. Luo, Y.; Kay, M. S.; Baldwin, R. L. *Nat. Struct. Biol.* **1997**, 4, 925.
  47. Martinez, J. S.; Pisabaro, M. T.; Serrano, L. *Nat. Struct. Biol.* **1998**, 5, 720.
  48. Otzen, D. S.; Elmarsy, N.; Jackson, S. E.; Fersht, A. R. *Proc. Natl. Acad. Sci., USA* **1994**, 91, 10422.
  49. Plaxco, K. W.; Baker, D. *Proc. Natl. Acad. Sci., USA* **1998**, 95, 13591.
  50. Pagget, V.; Li, A.; Itzahaki, L. S.; Otzen, D. E.; Fersht, A. R. *J. Mol. Biol.* **1996**, 257, 430.
  51. Kuszewski, J.; Clore, G. M.; Gronenborn, A. M. *Protein Sci.* **1994**, 3, 1945.
  52. Denton, M. E.; Rothwart, D. M.; Scheraga, H. A. *Biochemistry* **1994**, 33, 11125.
  53. Oas, T. G.; Kim, P. S. *Nature* **1988**, 336, 42.
  54. Baldwin, R. L.; Roder, H. *Curr. Biol.* **1991**, 1, 218.
  55. Nozaki, Y.; Tanford, C. *J. Biol. Chem.* **1971**, 246, 2211.
  56. Ikai, A.; Tanford, C. *Nature* **1971**, 230, 100.
  57. Ikai, A.; Tanford, C. *J. Mol. Biol.* **1973**, 73, 145.
  58. Ikai, A.; Fish, W. W.; Tanford, C. *J. Mol. Biol.* **1973**, 73, 165.
  59. Tanford, C.; Aune, K. C.; Ikai, A. *J. Mol. Biol.* **1973**, 73, 185.
  60. Tanford, C. *Adv. Protein Chem.*, **1968**, 23, 121.
  61. Tsong, T. Y.; Baldwin, R. L.; Elson, E. L. *J. Proc. Natl. Acad. Sci., USA* **1971**, 68, 2712.
  62. Tsong, T. Y.; Baldwin, R. L.; Elson, E. L. *Proc. Natl. Acad. Sci., USA* **1972**, 69, 1809.
  63. Tsong, T. Y.; Baldwin, R. L.; McPhie, P.; Elson, E. L. *J. Mol. Biol.* **1972**, 63, 453.
  64. Tsong, T. Y. *Biochemistry* **1973**, 12, 2209.
  65. Brandts, J. P.; Halnoson, H. R.; Brenman, M. *Biochemistry* **1975**, 14, 4953.
  66. Lin, L. N.; Brandts, J. F. *Biochemistry* **1978**, 17, 4102.
  67. Crieghton, T. E. *Prog. Biophys. Mol. Biol.* **1978**, 33, 231.
  68. Crieghton, T. E. *Biochem. J.* **1990**, 270, 1.
  69. Van Mierlo, C. P. M.; Darby, N. J.; Keeler, J.; Neuhaus, D.; Creighton, T. E. *J. Mol. Biol.* **1993**, 229, 1125.
  70. Karplus, M.; Weaver, D. L. *Protein Sci.* **1994**, 3, 658.
  71. Karplus, M.; Shakhovich, E. in *Protein Folding*; Creighton, T. E., Ed.; Freeman Press: San Francisco, **1995**; p 127.
  72. Bhattacharya, R. P.; Sosnick, T. R. *Biochemistry* **1999**, 38, 2601.
  73. Jacob, M.; Geever, M.; Hoiltermann, G.; Schmid, F. X. *Nat. Struct. Biol.* **1999**, 6, 923.
  74. Jacob, M.; Schmid, F. X. *Biochemistry* **1999**, 38, 13773.
  75. Kim, P. S.; Baldwin, R. L. *Ann. Rev. Biochem.* **1982**, 51, 459.
  76. Baldwin, R. L. *Proc. Natl. Acad. Sci., USA* **1996**, 93, 2627. Ohgushi, M.; Wada, A. *FEBS Lett.* **1983**, 164, 21.
  77. Brems, D. N.; Baldwin, R. L. *Biochemistry* **1985**, 24, 1689. Kuwajima, K. *Proteins: Struct. Funct. Genet.* **1989**, 6, 87.
  78. Sali, A.; Shakhovich, E.; Karplus, M. *J. Mol. Biol.* **1994**, 235, 161.
  79. Barrick, D.; Baldwin, R. L. *Biochemistry* **1993**, 32, 3790.
  80. Evans, P. A.; Radford, S. E. *Curr. Opin. Struct. Biol.* **1994**, 4, 100.
  81. Jeng, M. F.; Englander, S. W.; Elove, G. A.; Wand, A. J.; Roder, H. *Biochemistry* **1990**, 29, 10433.
  82. Miranker, A.; Radford, S. E.; Karplus, M.; Dobson, C. M. *Nature* **1991**, 349, 633.
  83. Radford, S. E.; Dobson, C. M.; Evans, P. A. *Nature* **1992**, 358, 302.
  84. Dobson, C. M.; Sali, A.; Karplus, M. *Angew. Chem.* **1998**, 37, 868.
  85. Shao, X.; Mathews, C. R. *Biochemistry* **1998**, 37, 7850.
  86. Briggs, M. S.; Roder, H. *Proc. Natl. Acad. Sci., USA* **1992**, 89, 2017.
  87. Koide, S.; Dyson, H. J.; Wright, P. E. *Biochemistry* **1993**, 32, 12999.
  88. Lui, Z. P.; Rizo, J.; Gierasch, L. M. *Biochemistry* **1994**, 33, 134.
  89. Villegas, V.; Azuga, A.; Catasius, L.; Reverter, D.; Mateo, P. L.; Aviles, F. X.; Serrano, L. *Biochemistry* **1995**, 34, 15105-15110.
  90. Harrison, S. C.; Durbin, R. *Proc. Natl. Acad. Sci., USA*

- 1985, 82, 4028.
91. Wetlaufer, D. *Proc. Natl. Acad. Sci., USA* **1973**, *94*, 6170.
92. Go, N. *Ann. Rev. Biophys. Bioeng.* **1983**, *12*, 183.
93. Kanehisa, M. I.; Tsong, T. Y. *J. Mol. Biol.* **1978**, *124*, 177.
94. Kanehisa, M. I.; Tsong, T. Y. *J. Mol. Biol.* **1979**, *133*, 279.
95. Camacho, C. J.; Thirumalai, D. *Proteins: Struct. Funct. Genet.* **1995**, *22*, 27.
96. Govindarajan, S.; Goldstein, R. A. *Proteins: Struct. Funct. Genet.* **1995**, *22*, 413.
97. Dill, K. A.; Bromberg, S.; Yue, K.; Fiebig, K. M.; Yee, D. P.; Thomas, P. D.; Chan, H. S. *Protein Sci.* **1995**, *4*, 561.
98. Zwanzig, R.; Szabo, A.; Bagachi, B. *Natl. Acad. Sci., USA* **1992**, *89*, 20.
99. Leopold, P. E.; Montal, M.; Onuchic, J. N. *Proc. Natl. Acad. Sci., USA* **1992**, *89*, 8721.
100. Guo, Z.; Thirumalai, D. *Biopolymers* **1995**, *36*, 83.
101. Karplus, M. *Folding and Design* **1997**, *2*, 69.
102. Irback, A.; Peter son, C.; Putthast, F. *Proc. Natl. Acad. Sci., USA* **1996**, *93*, 9533.
103. McCammon, J. A. *Proc. Natl. Acad. Sci., USA* **1993**, *93*, 11426-11427.
104. Jackson, S. E.; ElMasry, N.; Fersht, A. R. *Biochemistry* **1993**, *32*, 11270.
105. Jackson, S. E.; Fersht, A. R. *Biochemistry* **1991**, *30*, 10428.
106. Matouscheck, A.; Fersht, A. R. *Proc. Natl. Acad. Sci., USA* **1989**, *90*, 7814.
107. Fersht, A. R.; Matouschek, A.; Serrano, L. *J. Mol. Biol.* **1992**, *224*, 771.
108. Viguera, A. R.; Marti nez, J. C.; Filionov, V. V.; Mateo, R. I.; Serrano, L. *Biochemistry* **1994**, *33*, 2142.
109. Prieto, J.; Williams, M.; Jimenez, M. A.; Rico, M.; Serrano, L. *J. Mol. Biol.* **1997**, *268*, 760.
110. Chan, H. S.; Dill, K. A. *J. Chem. Phys.* **1994**, *100*, 9238.
111. Onuchic, J. N.; Wolynes, P. G.; Luthey-Schlt en, Z.; Socc i, N. D. *Proc. Natl. Acad. Sci., USA* **1995**, *92*, 3626.
112. Shakhnovich, E. I.; Gutin, A. M. *Proc. Natl. Acad. Sci., USA* **1993**, *90*, 7195.
113. Wolynes, P. G.; Onuchich, J. N.; Thirumalai, E. *Science* **1995**, *267*, 1619.
114. Abkevich, V. I.; Gutin, A. M.; Shakhnovich, E. I. *J. Chem. Phys.* **1994**, *101*, 6052.
115. Abkevich, V. J.; Gutin, A. M.; Shakhnovich, E. *Biochemistry* **1994**, *33*, 10026.
116. Chan, H. S. *Nature* **1995**, *373*, 664.
117. Jennings, P. A.; Wright, P. E. *Science* **1993**, *262*, 8, 892.
118. Var ley, P.; Gronenborn, A. M.; Christensen, H.; Wingfield, P. T.; Pain, R. H.; Clore, G. M. *Science* **1993**, *260*, 1110.
119. Itzhaki, L. S.; Evans, P. A.; Dobson, C. M.; Radford, S. E. *Biochemistry* **1994**, *33*, 5212.
120. Kuwajima, K.; Yamaya, H.; Sugai, S. *J. Mol. Biol.* **1996**, *264*, 806.
121. Thomas, P. D.; Dill, K. A. *Protein Sci.* **1993**, *2*, 2050.
122. Bychova, V. E.; Ptitsyn, O. B. *Chemtracts- Biochem. Molec. Biol.* **1993**, *4*, 133.
123. Baldwin, R. L.; Creighton, T. E. in *Protein Folding*; Jaenicke, R, Ed.; Elsevier: Amsterdam, **1980**; p 217.
124. Englander, S. W.; Mayne, L. *Ann. Rev. Biophys. Biomol. Struct.* **1992**, *21*, 243.
125. Mullins, L. S.; Pace, C. N.; Raushel, F. M. *Biochemistry* **1993**, *32*, 6152.
126. Bai, Y.; Milne, J. S.; Mayne, L.; Englander, S. W. *Proteins: Struct. Funct. Genet.* **1993**, *17*, 75.
127. Jamin, M.; Baldwin, R. L. *Nat. Struct. Biol.* **1996**, *3*, 613.
128. Balbach, J.; Baldwin, R. L. *Nat. Struct. Biol.* **1995**, *2*, 865.
129. Hoetzi, S. D.; Frieden, C. *Proc. Natl. Acad. Sci., USA* **1995**, *92*, 9318.
130. Hvidt, A.; Nielsen, S. O. *Adv. Protein Chem.* **1966**, *21*, 287.
131. Jacobs, M. P.; Fox, R. O. *Proc. Natl. Acad. Sci., USA* **1994**, *91*, 449.
132. Farrow, N. A.; Zhang, O.; Forman-Kay, J. D.; Kay, L.E. *Biochemistry* **1995**, *34*, 868.
133. Khorasanizadeh, S.; Peters, I. D.; Roder, H. *Nat. Struct. Biol.* **1996**, *3*, 193.
134. Sosnick, T. R.; Mayne, L.; Hilter, R.; Englander, S. W. *Nat. Struct. Biol.* **1994**, *1*, 149.
135. Roder, H.; Colton, W. *Curr. Opin. Struct. Biol.* **1997**, *7*, 15.
136. Gutin, A. M.; Abkevich, V. I.; Shakhnovich, E. *Biochemistry* **1995**, *34*, 3066.
137. Camacho, C. J.; Thirumalai, D. *Proc. Natl. Acad. Sci., USA* **1993**, 6369.
138. Wolynes, P. G. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 6170.
139. Dill, K. A.; Chan, H. S. *Nature Struct. Biol.* **1997**, *4*, 10.
140. Baldwin, R. L. *J. Biomolecular NMR* **1995**, *5*, 103.
141. Chan, H. S.; Dill, K. E. *Proteins: Struct. Funct. Genet.* **1998**, *30*, 2.
142. Huang, G. S.; Oas, T. G. *Biochemistry* **1995**, *34*, 3884.
143. Finkelstein, A. V. *Protein: Struct. Funct. Genet.* **1991**, *9*, 23.
144. Notting, B.; Golbik, R.; Niera, J. L.; Soler-Gonzales, A. O.; Schreiber, G.; Fersht, A. R. *Proc. Natl. Acad. Sci., USA* **1997**, *94*, 825.
145. Eaton, W. A.; Munoz, V.; Thompson, P. A.; Chanjc, C. K.; Hofrichter, J. *Curr. Opin. Struct. Biol.* **1997**, *7*, 10.

146. Baldwin, R. L. *Folding and Design* **1996**, 1, 1.
147. Riddle, O. S.; Santigo, J. V.; Bray Hal, S. T.; Doshi, N.; Grantacharova, V.; Yi, O.; Baker, D. *Nat. Struc. Biol.* **1977**, 4, 805.
148. Kragelund, B. B.; Robinson, C. V.; Knudsen, J.; Dobson, C. M.; Poulsen, F. M. *Biochemistry* **1997**, 34, 7217.
149. Parker, M.; Dempsey, C.; Kirch, M.; Clarke, A. R. *Biochemistry* **1997**, 36, 13395.
150. Clarke, J.; Harnill, S. J.; Johnson, C. M. *J. Mol. Biol.* **1997**, 270, 771.
151. Matouschek, A.; Kellis, J. T. Jr.; Serrano, L.; Fersht, A. R. *Nature* **1989**, 342, 122.
152. Matouschek, A.; Kellis, J. T. Jr.; Serrano, L.; Bycroft, M.; Fersht, A. R. *Nature* **1990**, 346, 440.
153. Matouschek, A.; Fersht, A. R. *Methods Enzymol.* **1991**, 202, 81.
154. Matouschek, A.; Serrano, L.; Fersht, A. R. *J. Mol. Biol.* **1992**, 224, 819.
155. Serrano, L.; Matouschek, A.; Fersht, A. R. *J. Mol. Biol.* **1992**, 224, 805.
156. Serrano, L. *Biochemistry* **1995**, 34, 15105.
157. Ropson, I. J.; Frieden, C. *Proc. Natl. Acad. Sci., USA* **1992**, 89, 7222.
158. Laurents, D. V.; Bruix, M.; Jamin, M.; Baldwin, R. L. *J. Mol. Biol.* **1998**, 275, 669.
159. Jaenicke, R. *Prog. Biophys. Mol. Biol.* **1987**, 49, 117.
160. Baldwin, R. L.; Rose, G. D. *Trends Biochem. Sci.* **1999**, 24, 26.
161. Shiraki, K.; Nishikawa, K.; Goto, Y. *J. Mol. Biol.* **1995**, 245, 180.
162. Baldwin, R. L. *Biophys. Chem.* **1995**, 127.
163. Creamer, T. R.; Rose, G. D. *ProteinSci.* **1995**, 1305.
164. Hughson, F. M.; Wright, P. E.; Baldwin, R. L. *Science* **1990**, 249, 1544.
165. Dyson, H. J.; Wright, P. E. *Curr. Opin. Struc. Biol.* **1993**, 3, 60.
166. Jiang, N.; Frieden, C. *Biochemistry* **1993**, 32, 11015.
167. Lindgren, M.; Svensson, M.; Freskgard, P. O.; Carlsson, U.; Jonasson, P. *Biochemistry* **1995**, 34, 657.
168. Gillespie, J. R.; Shortle, D. *J. Mol. Biol.* **1997**, 268, 170.
169. Roder, H.; Elove, G. A.; Englander, S. W. *Nature* **1988**, 335, 700.
170. Udgaonkar, J. B.; Baldwin, R. L. *Nature* **1988**, 335, 694.
171. Kiefhaber, T. *Proc. Natl. Acad. Sci., USA* **1995**, 92, 9029.
172. Blanco, F. J.; Jimenez, A.; Pineda, A.; Rico, M.; Santoro, J.; Nieto, J. L. *Biochemistry* **1994**, 33, 6004.
173. Minor, D. L. Jr.; Kim, P. S. *Nature* **1996**, 380, 730.
174. Kuroda, V.; Hamada, D.; Tanaka, T.; Goto, Y. *Folding and Design* **1996**, 1, 243.
175. Chaffote, A. F.; Guilou, Y.; Goldberg, M. E. *Biochemistry* **1992**, 31, 9694.
176. Sivaraman, T.; Kumar, T. K. S.; Lin, W. Y.; Chang, D. K.; Yu, C. *J. Biol. Chem.* **1998**.
177. Alexandrescu, A. T.; Evans, P. A.; Pitkeathly, M.; Baum, J.; Dobson, C. M. *Biochemistry* **1993**, 32, 1707.
178. Schonbrunner, N.; Wey, J.; Engels, J.; Georg, L. J.; Kiefhaber, T. *J. Mol. Biol.* **1996**, 260, 432.
179. Kumar, T. K. S.; Jayaraman, G.; Lee, C. S.; Sivaraman, T.; Lin, T. Y.; Yu, C. *Biochem. Biophys. Res. Commun.* **1995**, 207, 536.
180. Hamada, D.; Segawa, S.; Goto, Y. *Nat. Struc. Biol.* **1996**, 3, 868.
181. Oliverberg, M.; Tan, Y. J.; Fersht, A. R. *Proc. Natl. Acad. Sci., USA* **1995**, 92, 8926.
182. Srinivasan, R.; Rose, G. D. *Protein: Struct. Funct. Genet.* **1995**, 22, 81.
183. Aune, K. C.; Salahuddin, A.; Zarlengo, M. H.; Tanford, C. *J. Biol. Chem.* **1967**, 242, 4486.
184. Privalov, P. L. *Adv. Protein Chem.* **1979**, 33, 167.
185. Privalov, P. L. *Adv. Protein Chem.* **1982**, 35, 1.
186. Privalov, P. L.; Makhatadze, G. I. *J. Mol. Biol.* **1990**, 213, 385.
187. Ptitsyn, O. B. *Adv. Protein Chem.* **1995**, 47, 83.
188. Fiebig, K. M.; Schwalbe, H.; Buck, M.; Smith, L. J.; Dobson, C. M. *J. Phys. Chem.* **1996**, 100, 2661.
189. Neri, D.; Billeter, M.; Wider, G.; Wuthrich, K. *Science* **1992**, 257, 1559.
190. Alexandrescu, A. T.; Abeygunawardana, C.; Shortle, D. *Biochemistry* **1994**, 33, 1063.
191. Buck, M.; Radford, S. E.; Dobson, C. M. *Biochemistry* **1993**, 32, 669.
192. Stocman, B. J.; Euvrard, A.; Schahill, T. A. *J. Biomolec. NMR.* **1993**, 3, 285.
193. Redfield, C.; Smith, R. A. G.; Dobson, C. M. *Nat. Struc. Biol.* **1994**, 1, 23.
194. Feng, Y.; Sligar, S. G.; Wand, A. J. *Nat. Struc. Biol.* **1994**, 1, 30.
195. Bychkova, V. E.; Ptitsyn, O. B. *Biofizika (USSR)* **1992**, 23, 419.
196. Harding, M. M.; Williams, D. H.; Woolfson, D. N. *Biochemistry* **1991**, 30, 3120.
197. Izumi, Y.; Miyake, Y.; Kuwajima, K.; Sugai, S.; Inoue, K.; Izumi, M.; Katano, S. *Physica* **1983**, 12013, 444.
198. Matouschek, A.; Fersht, A. R. *Proc. Natl. Acad. Sci., USA* **1989**, 90, 7814.
199. Baum, J.; Dobson, C. M.; Evans, D. A.; Hanly, C. *Biochemistry* **1989**, 28, 7.
200. Jaganadham, M. V.; Balasubramanian, D. *FEBS Lett.* **1985**, 188, 326.
201. Bushnell, G. V.; Louie, G. V.; Brayer, G. D. *J. Mol. Biol.* **1990**, 214, 585.



202. Kumar, T. K. S.; Subbaiah, V.; Ramakrishna, T.; Pandit, M. W. *J. Biol Chem.* **1994**, *269*, 12620.
203. Dolgikh, D. A.; Abaturov, L. V.; Bolotina, I. A.; Brazhnikov, E. V.; Bychkova, V. E.; Bushuev, V. N.; Gilmanshin, R. I.; Lebedev, R.; Yu, O.; Semistnov, G. V.; Tiktopulo, E. I.; Ptitsyn, O. B. *Eur. Biophys. J.* **1985**, *13*, 119.
204. Timchenko, A. A.; Dolgikh, D. A.; Damaschun, H.; Damaschun, G. *Studia Biophys.* **1986**, *112*, 201.
205. Damaschun, G.; Gernat, C.; Damaschun, H.; Bychkova, V. E.; Ptitsyn, O. B. *Int. J. Biol. Macromol.* **1985**, *8*, 226.
206. Rodionova, N. A.; Semisotonov, G. V.; Kutysenko, V. P.; Rodionova, N. A.; Semistonov, G. V.; Kutysenko, V. P.; Uversky, V. N.; Bolotina, I. A.; Bychkova, V. E.; Ptitsyn, O. B. *Mol. Biol. (USSR)* **1989**, *23*, 683.
207. Gilmanshin, R. I.; Dolgikh, D. A.; Ptitsyn, O. B.; Finkelstein, A. V.; Shakhnovich, E. I. *Biofizika (USSR)* **1982**, *27*, 1005.
208. Dolgikh, D. A.; Gilmanshin, R. I.; Brazhnikov, E. V.; Bychkova, V. E.; Semisotnov, G. V.; Venyaminov, S.; Yu, O.; Ptitsyn, O. B. *FEBS Lett.* **1981**, *136*, 311.
209. Dolgikh, D. A.; Kolomiets, A. P.; Bolotina, I. A.; Ptitsyn, O. B. *FEBS Lett.* **1984**, *165*, 88.
210. Jeng, M. F.; Englander, S. W. *J. Mol. Biol.* **1991**, *221*, 1045.
211. Logan, T. M.; Therault, D.; Fesik, S. W. *J. Mol. Biol.* **1994**, *236*, 637.
212. Nelson, J. W.; Kallenbalch, N. R. *Biochemistry* **1989**, *28*, 5256.
213. Lehman, S.; Tuls, J. L.; Lund, M. *Biochemistry* **1990**, *29*, 5590.
214. Semisotonov, G. V.; Kutysenko, V. P.; Ptitsyn, O. B. *Mol. Biol. (USSR)* **1989**, *23*, 808.
215. Semisotnov, G. V.; Rodionova, N. A.; Kubysenko, V. P.; Ebert, B.; Blanco, J.; Ptitsyn, O. B. *FEBS Lett.* **1987**, *224*, 9.
216. Wong, K. P.; Hamlin, L. M. *Biochemistry* **1974**, *13*, 2678.
217. Finkelstein, A. V.; Badretdinov, A.; Ya, C.; Ptitsyn, O. B. *Prog. Biophys. Mol. Biol.* **1987**, *50*, 171.
218. Shakhnovich, E. I.; Finkelstein, A. V. *Biopolymers* **1989**, *28*, 1667.
219. Merril, A. R.; Cohen, F. S.; Cramer, W. A. *Biochemistry* **1990**, *29*, 5829.
220. Wang, T.; Chan, R.; Kallenbalch, N. R. *Proteins* **1998**, *30*, 435.
221. Pfeil, W.; Bychkova, V. E.; Ptitsyn, O. B. *FEBS Lett.* **1986**, *198*, 287.
222. Bychkova, V. E.; Pain, R. H.; Ptitsyn, O. B. *FEBS Lett.* **1983**, *238*, 231.
223. Gast, K.; Zirwer, D.; Welfle, H.; Bychkova, V. E.; Ptitsyn, O. B. *Int. J. Biol. Macromol.* **1986**, *8*, 231.
224. Ptitsyn, O. B. in *Protein Folding*; Creighton, T. E., Ed.; Freeman: New York, NY, **1992**; pp 243-300.
225. Schatz, G. *Protein Sci.* **1983**, *2*, 141.
226. Linsay, L. A.; Glover, J. G. *Biochem. J.* **1992**, *284*, 609.
227. Van der Goot, F. G.; Gonzalez-Munoz, J. M.; Lakey, J. H.; Pattus, F. *Nature* **1991**, *354*, 408.
228. Kagan, B. I.; Baldwin, R. L.; Munoz, D.; Wisniewski, B. J. *Science* **1992**, *255*, 1427.
229. Martin, J.; Langer, T.; Boteva, R.; Schramel, A.; Horwich, A. I.; Hartl, F. V. *Nature* **1991**, *352*, 36.
230. Chaing, H. L.; Terlecky, S. R.; Plant, C. P.; Dice, J. F. *Science* **1989**, *246*, 382.
231. Cohen, F. E. *J. Mol. Biol.* **1999**, *293*, 313.
232. Harrison, P. M.; Chan, H. S.; Prusiner, S. B.; Cohen, F. E. *J. Mol. Biol.* **1999**, *286*, 593.
233. Baldwin, M. A.; James, T. L.; Cohen, F. E.; Prusiner, S. B. *Biochem. Soc. Trans.* **1998**, *26*, 481.
234. Cohen, F. E.; Prusiner, S. B. *Ann. Rev. Biochem.* **1998**, *67*, 793.
235. Baum, J.; Brodsky, B. *Curr. Opin. Str. Biol.* **9**, 122.
236. Balbach, J.; Seip, S.; Kessler, H.; Scharf, M.; Kashani-Poor, N.; Engels, J. W. *Proteins* **1998**, *33*, 285.
237. Schonbrunner, N.; Pappenberger, G.; Scharf, M.; Engels, J.; Kiefhaber, T. *Biochemistry* **1997**, *36*, 9057.
238. Schonbrunner, N.; Koller, K. P.; Kiefhaber, T. *J. Mol. Biol.* **1996**, *268*, 526.
239. Viguera, A. R.; Serrano, L.; Wilmanns, M. *Nat. Struc. Biol.* **1996**, *3*, 874.
240. Pickford, A. R.; Potts, J. R.; Bright, J. R.; Phan, I.; Campbell, I. D. *Structure* **1997**, *5*, 359.
241. Spitzfaden, C.; Grant, R. P.; Mardon, H. J.; Campbell, I. D. *J. Mol. Biol.* **1997**, *265*, 565.
242. Plaxco, K. W.; Morton, C. J.; Gujjarro, J.; Pitkeathly, M.; Camp trell, I. D.; Dobson, C. M. *Biochemistry* **1988**, *37*, 438.
243. Martensson, L. G.; Jonsson, B. H. *Biochemistry* **1993**, *32*, 224. Plaxco, K. W.; Spitzfaden, C.; Campbell, J. D.; Dobson, C. M. *J. Mol. Biol.* **1997**, *270*, 763.
244. Plaxco, K. W.; Spitzfaden, C.; Campbell, I. P.; Dobson, C. M. *Proc. Natl. Acad. Sci.* **1996**, *93*, 10703.
245. Martensson, L. C.; Jonsson, B. H. *Biophys. J.* **1995**, *69*, 202.
246. Hilkansson, K.; Carlsson, M.; Svensson, L. A.; Liljas, A. *J. Mol. Biol.* **1992**, *227*, 1192.
247. Svensson, M.; Jonasson, P.; Freskgard, P. O.; Jonsson, B. H.; Lindgren, M.; Martensson, L. G.; Gentile, M.; Boren, K.; Carlsson, U. *Biochemistry* **1995**, *34*, 8606.
248. Oversky, V. N.; Ptitsyn, O. B. *J. Mol. Biol.* **1996**, *255*, 215.
249. Aronsson, G.; Martensson, L. G.; Carlsson, U.; Jonsson,

- B. H. *Biochemistry* **1995**, *34*, 2153.
250. Andersson, D.; Freskgard, P. O.; Jonsson, B. H.; Carlsson, U. *Biochemistry* **1997**, *36*, 4623.
251. Schlinder, T.; Herrler, M.; Marahiel, M. A.; Schmid, F. X. *Nat. Struct. Biol.* **1995**, *2*, 663.
252. Schrnchel, A.; Welschek, R.; Czish, M.; Herrler, M.; Willimsky, G.; Graumann, P.; Marahiel, M. A.; Holak, T. A. *Nature* **1993**, *364*, 169.
253. Schindler, T.; Schmid, F. X. *Biochemistry* **1996**, *35*, 16833.
254. Schindler, T.; Herrler, M.; Marahiel, M. A.; Schmid, F. X. *Nat. Struct. Biol.* **1995**, *2*, 668.
255. Jacob, M.; Holtermann, G.; Peri, D.; Reinstein, J.; Schindler, T.; Geeves, M.; Schmid, F. X. *Biochemistry* **1999**, *38*, 2882.
256. Peri, D.; Welker, C.; Schindler, T.; Schroder, K.; Marahiel, M. A.; Jaenicke, R.; Schmid, F. X. *Nat. Struct. Biol.* **1995**, *5*, 229.
257. Schlinder, T.; Graumann, P. L.; Perl, D.; Ma, S.; Schmid, F. X.; Marahiel, M. A. *J. Biol. Chem.* **1999**, *274*, 3407.
258. Thompson, J.; Brat, J.; Banazak, L. *J. Mol. Biol.* **1995**, *252*, 433.
259. Banazak, L.; Winter, N.; Xu, Z.; Bernoihr, D. A.; Cowan, S.; Jones, T. A. *Adv. Prot. Chem.* **1994**, *45*, 89.
260. Carlsson, U.; Jonsson, B. H. *Curr. Opin. Struct. Biol.* **1995**, *5*, 482.
261. Liu, Z. P.; Rizo, J.; Gierasch, L. M. *Biochemistry* **1994**, *33*, 134.
262. Clark, P. L.; Liu, Z. P.; Zhang, J.; Gierasch, L. U. *Protein Sci.* **1996**, *5*, 1108.
263. Clark, P. L.; Liu, Z. P.; Rizo, J.; Gierasch, L. M. *Nat. Struct. Biol.* **1997**, *4*, 883.
264. Auron, P. E.; Webb, A. C.; Rosenwasser, L. J.; Mucci, S. F.; Rich, A.; Wolff, S. M.; Dinavello, C. A. *Proc. Natl. Acad. Sci., USA* **1984**, *81*, 7907.
265. Priestle, J. P.; Schar, H. P.; Grutter, M. G. *Proc. Natl. Acad. Sci., USA* **1989**, *86*, 9667.
266. March, C. J.; Mosley, B.; Larsen, A.; Cerretti, D. P.; Braedt, G.; Price, V.; Gillis, S.; Henney, C. Z.; Kvonheim, S. R.; Grabstein, K.; Conlon, P. J.; Hopp, T. P.; Cosman, D. *Nature* **1985**, *315*, 640.
267. Chrnyk, B. A.; Evans, J.; Lilliquist, J.; Young, P.; Wetzel, R. *J. Biol. Chem.* **1993**, *268*, 18053.
268. Craig, S.; Schmeisser, V.; Wingfield, P.; Pain, R. H. *Biochemistry* **1987**, *268*, 18053.
269. Varley, P.; Gronenborn, A. M.; Christensen, H.; Wingfield, P. T.; Pain, R. H.; Clore, G. M. *Science* **1993**, *260*, 1110.
270. Gronenborn, A. M.; Clore, G. M. *Science* **1994**, *263*, 536.
271. Dill, K. A.; Fiebig, K. M.; Chan, H. S. *Proc. Natl. Acad. Sci., USA* **1993**, *90*, 1942.
272. Heidary, D. K.; Gross, L. A.; Roy, M.; Jennings, P. A. *Nat. Struct. Biol.* **1997**, *4*, 725.
273. Oberg, K.; Chrnyk, B. A.; Wetzel, R.; Fink, A. L. *Biochemistry* **1994**, *33*, 2628.
274. Hazes, B.; Hol, W. G. J. *Proteins: Struct. Funct. Genet.* **1992**, *12*, 278.
275. Williams, P. A.; Fulop, V.; Leung, Y. C.; Chan, C.; Moir, J. W. B.; Howlett, G.; Ferguson, S. J.; Radford, S. E.; Hajdu, J. *Nat. Struct. Biol.* **1995**, *2*, 975.
276. Capaldi, A. P.; Fergusson, S. J.; Radford, S. E. *J. Mol. Biol.* **1999**, *286*, 1621.
277. Capaldi, A. P.; Radford, S. E. *Curr. Opin. Struct. Biol.* **1998**, *8*, 86.
278. Kumar, T. K. S.; Sivaraman, T.; Yu, C. in *Natural and Synthetic Toxins: Biological Implications*; Oxford University Press, **1999** (in Press).
279. Kumar, T. K. S.; Pandian, S.; Jayaraman, G.; Peng, H. J.; Yu, C. *Proc. Natl. Acad. Sci., ROC* **1999**, *23*, 1.
280. Kumar, T. K. S.; Pandian, S.; Srisailam, S.; Yu, C. *J. Toxicol. Toxin. Rev.* **1998**, *17*, 183.
281. Kumar, T. K. S.; Jayaraman, G.; Lee, C. S.; Arunkumar, A. I.; Sivaraman, T.; Samuel, D.; Yu, C. *J. Biomolec. Struct. Dyn.* **1997**, *15*, 431.
282. Sivaraman, T.; Kumar, T. K. S.; Yang, P. W.; Yu, C. *Toxicon* **1997**, *35*, 1367.
283. Kumar, T. K. S.; Lee, C. S.; Yu, C. in *Natural Toxins*; Singh, B. R.; Tu, A.T., Eds; Plenum press: New York, **1996**; p 115.
284. Bhaskaran, R.; Huang, C. C.; Chang, D. K.; Yu, C. *J. Mol. Biol.* **1994**, *235*, 1291.
285. Kumar, T. K. S.; Yang, P. W.; Lin, S. H.; Wu, C. Y.; Lei, B.; Lo, S. J.; Tu, S. C.; Yu, C. *Biochem. Biophys. Res. Commun.* **1996**, *219*, 450.
286. Sivaraman, T.; Kumar, T. K. S.; Yu, C. *Biochemistry* **1999**, *38*, 9899.
287. Arunkumar, A. I.; Kumar, T. K. S.; Yu, C. *Biochim. Biophys. Acta.* **1997**, *1338*, 69.
288. Jayaraman, G.; Kumar, T. K. S.; Arunkumar, A. I.; Yu, C. *Biochem. Biophys. Res. Commun.* **1996**, *222*, 33.
289. Sivaraman, T.; Kumar, T. K. S.; Yu, C. *Int. J. Biol. Macromol.* **1996**, *19*, 235.
290. Arunkumar, A. I.; Kumar, T. K. S.; Jayaraman, G.; Samuel, D.; Yu, C. *J. Biomolec. Struct. Dyn.* **1996**, *14*, 381.
291. Kumar, T. K. S.; Jayaraman, G.; Lin, W. Y.; Yu, C. *Biochim. Biophys. Acta.* **1996**, *1294*, 103.
292. Shiraki, K.; Nishikawa, K.; Goto, Y. *J. Mol. Biol.* **1995**, *245*, 180.
293. Sivaraman, T.; Kumar, T. K. S.; Hung, K. W.; Yu, C. *J. Protein Chem.* **1999**, *18*, 481.
294. Kumar, T. K. S.; Jayaraman, G.; Lee, C. S.; Sivaraman,

- T.; Lin, W. Y.; Yu, C. *Biochem. Biophys. Res. Commun.* **1995**, *207*, 536.
295. Chang, J. Y.; Kumar, T. K. S.; Yu, C. *Biochemistry* **1998**, *37*, 6745.
296. Sivaraman, T.; Kumar, T. K. S.; Yu, C. *Biochem. Mol. Biol. Int.* **1998**, *44*, 29.
297. Sivaraman, T.; Kumar, T. K. S.; Jayaraman, G.; Yu, C. *Biochem. J.* **1997**, *321*, 457.
298. Jayaraman, G.; Kumar, T. K. S.; Sivaraman, T.; Lin, W. Y.; Chang, D. K.; Yu, C. *Int. J. Biol. Macromol.* **1996**, *18*, 303.
299. Sivaraman, T.; Kumar, T. K. S.; Chang, D. K.; Lin, W. Y.; Yu, C. *J. Biol. Chem.* **1998**, *273*, 10181.
300. Yu, C.; Bhaskaran, R.; Chuang, L. C.; Yang, C. C. *Biochemistry* **1993**, *32*, 2131.
301. Sivaraman, T.; Kumar, T. K. S.; Tu, Y. T.; Wang, W.; Lin, W. Y.; Chen, H. M.; Yu, C. *Biochem. Biophys. Res. Commun.* **1999**, *208*, 284.
302. DiGabriele, A. D.; Lay, I.; Chen, D. I.; Svahn, C. M.; Jaye, M.; Schleissinger, J.; Hendrickson, W. A. *Nature* **1998**, *393*, 812.
303. Zhu, X.; Hsu, B. T.; Rees, D. C. *Structure* **1993**, *1*, 27.
304. Zhu, X.; Komiya, H.; Chirno, A.; Fahamo, S.; Fox, G. M.; Arakawa, T.; Hsu, B. T.; Rees, D. C. *Science* **1991**, *253*, 90.
305. Habazetti, J.; Gondol, D.; Witscheck, R.; Otlewski, J.; Schleicher, M.; Holak, T. A. *Nature* **1992**, 855.
306. Burke, C. J.; Volkin, D. B.; Mach, H.; Middaugh, C. R. *Biochemistry* **1993**, *32*, 6419.
307. Sanz, J. M.; Gallego, G. *Eur. J. Biochem.* **1997**, *246*, 328.
308. Mach, H.; Middaugh, C. R. *Arch. Biochem. Biophys.* **1994**, *309*, 36.
309. Dabora, J. M.; Sanyal, G.; Middaugh, C. R. *J. Biol. Chem.* **1991**, *256*, 23637.
310. Ragona, L.; Pusteria, F.; Zelta, L.; Monaco, H.; Molinari, H. *Folding & Design* **1997**, *2*, 281.
311. Aakerston, B.; Loegerberg, L. *Trend, Biochem. Sci.* **1990**, *15*, 240.
312. Flower, D. R. *J. Mol. Recognit.* **1997**, *8*, 185.
313. Brownlow, S.; Sawyer, L. *Structure* **1997**, *5*, 481.
314. Molinary, H.; Monaco, H. L. *FEBS Lett.* **1996**, *381*, 237.
315. Fujiwara, K.; Arai, M.; Shimizu, A.; Ikeguchi, M.; Kuwajima, K.; Sugai, S. *Biochemistry* **1999**, *38*, 4455.
316. Dufour, F.; Hoa, G. H. B.; Haertle, T. *Biochim. Biophys. Acta.* **1994**, *1206*, 166.
317. Dufour, F.; Genot, C.; Haertle, T. *Biochim. Biophys. Acta.* **1994**, *1205*, 105.
318. Kuwajima, K.; Yamaha, H.; Miwa, S.; Sugai, S.; Nagamura, T. *FEBS Lett.* **1987**, *221*, 115.
319. Kuwajima, K.; Yamaha, H.; Sugai, S. *J. Mol. Biol.* **1996**, *264*, 806.
320. Hamada, D.; Kuroda, Y.; Tanaka, T.; Goto, Y. *J. Mol. Biol.* **1995**, *254*, 737.
321. Hamada, D.; Goto, Y. *J. Mol. Biol.* **1997**, *269*, 479.
322. Hamada, D.; Segawa, S.; Goto, Y. *J. Mol. Biol.* **1995**, *254*, 737.
323. Arai, M.; Ikura, T.; Semisotonov, G. V.; Kihara, H.; Amemiya, Y.; Kuwajima, K. *J. Mol. Biol.* **1998**, *275*, 149.
324. Subramanian, V.; Steel, D. G.; Gafni, A. *ProteinScience* **1996**, *5*, 2089.
325. Hattori, M.; Amefani, A.; Katakura, Y.; Shimizu, M.; Kamanigowa, S. *J. Mol. Biol.* **1993**, *268*, 22414.
326. Koch, C. A.; Anderson, D.; Moran, M. F.; Ellis, C.; Pawson, T. *Science* **1991**, *252*, 668.
327. Musacchio, A.; Noble, M. E. M.; Pautit, R.; Wierenga, R.; Saraste, M. *Nature* **1992**, *359*, 851.
328. Noble, M. E. M.; Musacchio, A.; Saraste, M.; Courteidge, S. A.; Wierenga, R. K. *EMBO. J* **1993**, *12*, 2617.
329. Yu, H.; Rosen, M. K.; Shin, T. B.; Siedel-Dugan, C.; Brugge, J. S.; Schreiber, S. L. *Science*, **1992**, *258*, 1665.
330. Viguera, A. R.; Martinez, J. C.; Filimonov, V. V.; Maten, P. L.; Serrano, L. *Biochemistry* **1994**, *33*, 2142.
331. Farrow, N. A.; Zhang, O.; Forman-Kay, J. D.; Kay, L. E. *Biochemistry* **1995**, *34*, 868.
332. Zhang, O.; Forman-Kay, J. D. *Biochemistry* **1995**, *34*, 6784.
333. Zhang, O.; Forman-Kay, J. D. *Biochemistry* **1997**, *36*, 3959.
334. Blanco, F.; Serrano, L.; Forman-Kay, J. D. *J. Mol. Biol.* **1998**, *284*, 1153.
335. Mok, Y. K.; Kay, C. M.; Kay, L. E.; Forman-Kay, J. D. *J. Mol. Biol.* **1999**, *289*, 619.
336. Viguera, A. R.; Blanco, F. J.; Serrano, L. *J. Mol. Biol.* **1995**, *247*, 670.
337. Viguera, A. R.; Serrano, L.; Wilmanns, M. *Nat. Struct. Biol.* **1996**, *3*, 874.
338. Viguera, A. R.; Serrano, L. *Nat. Struct. Biol.* **1997**, *4*, 939.
339. Prieto, J.; Wilmanns, M.; Jimenez, M. A.; Rico, M.; Serano, L. *J. Mol. Biol.* **1997**, *268*, 760.
340. Guijaro, J. L.; Mor-ton, C. J.; Plaxco, K. W.; Campbell, I. D.; Dobson, C. M. *J. Mol. Biol.* **1998**, *276*, 657.
341. Tsai, J.; Levitt, M.; Baker, D. J. *J. Mol. Biol.* **1999**, *291*, 215.

