10 Protein Folding

A. Szilágyi · J. Kardos · S. Osváth · L. Barna · P. Závodszky

Abstract: Since Anfinsen's famous experiments in the 1960s, it has been known that the complex three-dimensional structure of protein molecules is encoded in their amino acid sequences, and the chains autonomously fold under proper conditions. Cracking this code, which is sometimes called "the second part of the genetic code," has been one of the greatest challenges of molecular biology. Although a full understanding of how proteins fold remains elusive, theoretical and experimental studies of protein folding have come a long way since Anfinsen's findings. In the living cell, folding occurs in a complex and crowded environment, often involving helper proteins, and in some cases it can go awry: the protein can misfold, aggregate, or form amyloid fibers. It is increasingly recognized that misfolded proteins and amyloid formation are the root cause of a number of serious illnesses including several neurodegenerative diseases. Therefore, the study of protein folding remains a key area of biomedical research.

List of Abbreviations: AFM, atomic force microscopy; CCP, chaperonin-containing TCP-1; CIDNP, chemically induced nuclear polarization; DMD, discrete molecular dynamics; EM, electron microscopy; ESIMS, electrospray ionization mass spectrometry; FIS, factor for inversion stimulation; FMN, flavin mononucleotide; NOE, nuclear Overhauser effect; TF, trigger factor; TCP-1, tailless complex polypeptide-1

1 Protein Structure and Its Physical Basis

The function of a protein can only be interpreted from its structure. The nervous system is a network of cells, and the peculiar functional properties of these cells can be derived from the properties and interactions of their proteins. Proteins are involved in all stages of neural activity. Those embedded wholly or partly in membranes regulate the transport of ions and molecules as a means of signal exchange with other cells and the external medium. Some of them have enzymatic functions to catalyze the chemical processes essential for function. The diverse and highly specific function of proteins is a consequence of their sophisticated, individual surface pattern regarding shape, charge, and hydrophobicity. The surface pattern is a consequence of the unique three-dimensional structure of the polypeptide chain. Proteins are linear polymers with nonrepetitive, specific covalent structure. The covalent structure is determined by the order of amino acids in which they are linked together. Since Anfinsen's famous experiments (1973) in the 1960s, it has been believed and today generally accepted that folding and the resulting native structure of proteins are autonomously governed and determined by the amino acid sequence of a particular protein and its natural solvent environment.

1.1 Physical Forces and Principles Underlying Protein Folding and Structure

A linear polypeptide chain is autonomously organized into a space-filling, compact, and well-defined three-dimensional structure. In a globular protein, the internal core is mostly formed by hydrophobic amino acid residues, held together by van der Waals forces, and the surface of the globule is formed by mostly charged and polar side chains. Proteins exist in this state of condensed matter while the specific conformation is largely determined by the flexibility of the polypeptide backbone and by the specific, consistent intermolecular interactions of the side chains.

The monomeric unit in a polypeptide chain is the peptide group. The sequence of amino acids is the primary structure of the protein. The C, O, N, and H atoms lie in the same plane; successive planes define angles ϕ and ψ. The conformation of a chain of n amino acids can be defined by $2n$ parameters. The restricted flexibility of the polypeptide chain is a major factor among those determining protein structure and folding.

The native conformation must be energetically stable. From a thermodynamic point of view, the free energy of a protein molecule is influenced by the following major contributions: (1) the hydrophobic effect, (2) the energy of hydrogen bonds, (3) the energy of electrostatic interactions, and (4) the conformational entropy due to the restricted motion of the main chain and the side chains.

The hydrophobic effect used to be explained as a primarily entropic effect arising from the rearrangement of hydrogen bonds between solvent molecules around an apolar solute. This hydration process is energetically unfavorable, and therefore drives apolar solutes together, thereby decreasing their solvent-exposed surface area. Today, the hydrophobic effect is usually viewed as a combined effect of hydration (an entropic effect) and van der Waals interactions between solute molecules (an enthalpic effect) (Makhatadze and Privalov, 1995). It is therefore entropic at low temperatures and enthalpic at high temperatures, which results in a complex temperature dependence of its strength (Schellman, 1997). Nevertheless, the hydrophobic force has long been considered as the major driving force of protein folding (Dill, 1990) as it leads to a rapid collapse of the polypeptide chain, thereby largely reducing the configurational space to explore. Without doubt, the hydrophobic interaction is also a major stabilizing force contributing to the thermodynamic stability of the folded state.

The role of hydrogen bonds in folding and stability used to be underestimated based on the argument that intramolecular hydrogen bonds can be replaced by hydrogen bonds between the protein and the solvent. After a number of mutational studies, however, hydrogen bonds have now been recognized as having a contribution to protein stability as important as the hydrophobic effect (Pace et al., 1996). This contribution was estimated to be 1.5 ± 1.0 kcal/mol per buried intramolecular hydrogen bond (Pace et al., 1996).

Electrostatic interactions such as ion pairs and salt bridges in proteins have been an area of active research (Kumar and Nussinov, 2002). While hydrogen bonds and hydrophobic forces are essentially nonspecific, electrostatic interactions are largely specific, and therefore play an important role in specifying the fold of a protein as well as in protein flexibility and function. Computational and experimental evidence shows that salt bridges can be stabilizing or destabilizing. On the other hand, genome-wide and structural comparisons of thermophilic and mesophilic proteins indicate that salt bridges may significantly contribute to the enhanced thermal stability of proteins from thermophilic organisms (Szilagyi and Zavodszky, 2000; Li et al., 2005; Razvi and Scholtz, 2006).

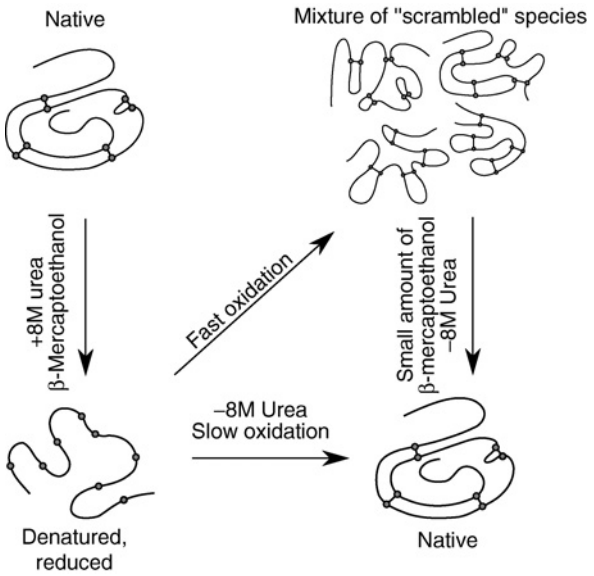
The major destabilizing contribution to the stability of the folded state is the conformational entropy of the polypeptide chain. Folding a long chain into a specific, compact structure obviously results in a significant entropy decrease. This is counterbalanced by the various intrachain interactions described above. The resulting overall stability of the protein (the free-energy difference between the folded and the unfolded state) is marginal, being on the order of 5–10 kcal/mol. This number is a small difference between huge stabilizing and destabilizing contributions. We qualitatively know that the hydrophobic effect and hydrogen bonds are the major stabilizing contributions and the conformational entropy is the major destabilizing one. However, due to the compensatory effects in the total energy balance, a quantitative prediction with respect to the significance of any specific type of interaction cannot be made with confidence (Jaenicke, 2000).

2 The Protein-Folding Problem

The "central dogma" of molecular biology states that the flow of sequential information from nucleic acid to protein is unidirectional: nucleic acid sequences encode the sequence of proteins but once translation occurs, information cannot flow back from protein to nucleic acid. A possible extension of the central dogma would be to add that the sequence of the protein "encodes" its three-dimensional structure. Indeed, this coding is sometimes called the "second half of the genetic code." Cracking this code would be equivalent to solving the "protein folding problem." To see what this problem is all about, let us look at Anfinsen's classic experiment (1973) (❯ *Figure 10-1*).

Ribonuclease, an enzyme with 124-amino acid residues, contains four disulfide bridges. Treatment of ribonuclease with 8 M urea in the presence of the reducing agent β-mercaptoethanol causes a complete unfolding of the ribonuclease molecule, yielding an essentially random form. In this process, the four disulfide bridges get cleaved, resulting in eight free SH groups. The enzymatic activity of the molecule is completely lost. Allowing the cysteines to reoxidize under denaturing conditions results in a mixture of "scrambled" species where the eight SH groups randomly pair to form four disulfide bridges (it is easy to

Figure 10-1
Scheme of Anfinsen's experiment. See text for details



calculate that there are 105 possible pairings). However, when urea is slowly removed by dialysis and a small amount of β-mercaptoethanol is added, disulfide interchange takes place and the mixture of "scrambled" ribonuclease molecules is eventually converted to a homogeneous product, which is fully enzymatically active and indistinguishable from native ribonuclease. This crucial experiment supports the idea known as the "thermodynamic hypothesis," which states that the three-dimensional structure of a native protein in its normal physiological state is the one in which the Gibbs free energy of the whole system is lowest. Therefore, in a given environment, the native conformation of the protein is fully determined by its amino acid sequence. How the information encoded in the sequence gets translated into the three-dimensional structure? This is the protein-folding problem.

Assuming that the polypeptide chain randomly samples all possible configurations, we can estimate the time required for a protein to fold. If each bond connecting two neighboring amino acids can have, say, three possible states then a protein of 101 amino acids could exist in $3^{100} = 5 \times 10^{47}$ different configurations. Only one of these configurations corresponds to the native state. Even if the protein is able to sample new configurations at a rate of 10^{13} per second, it will take 10^{27} years to try them all. Nevertheless, proteins do fold, and in a timescale of seconds or less. This contradiction was first pointed out by Cyrus Levinthal in 1969 (Levinthal, 1969) and has become known as "Levinthal's paradox." To resolve the paradox, Levinthal argued that the protein cannot fold by random search and there must be specific "folding pathways."

The concept of folding pathways motivated a large number of experimental studies aimed at finding the specific "folding intermediates" and also gave rise to a number of models describing the folding process. For example, the nucleation/growth model (Wetlaufer, 1973) tried to resolve Levinthal's paradox by assuming that the rate-limiting step of the folding process is a nucleation event, presumably the formation of smaller structural units, and once nucleation occurs the nuclei grow fast and the folding process rapidly completes. This model is not consistent with the large number of observations where folding intermediates were observed. According to the "diffusion–collision–adhesion model" (Karplus and Weaver, 1976), fluctuating microdomains (portions of secondary structure or hydrophobic clusters) move diffusively and repeatedly collide with each other. Collisions can lead to a coalescence of the microdomains into larger units (adhesion). The rate-limiting stage is assumed to be the diffusion process. This model is well supported by many experiments (Karplus and Weaver, 1994). The "framework model" (Baldwin, 1989) states that the

folding process is hierarchical, starting with the formation of the secondary structure elements, and the docking of the preformed substructures is the rate-limiting step. The "hydrophobic collapse model" (Dill, 1985) is based on the view that the hydrophobic effect is the main driving force of folding, and the process starts with a rapid collapse of the chain, followed by the formation of the secondary structure. In fact, whether hydrophobic collapse or secondary structure formation occurs first has remained a largely undecided issue even to this day. Finally, the "jigsaw puzzle model" (Harrison and Durbin, 1985) denied the necessity of a unique, directed folding pathway and stated that each protein molecule can follow a different route to the native structure, just like there are multiple ways to solve a jigsaw puzzle. This idea is actually consistent with a "new view" of protein folding, which gained popularity in the 1990s: the energy landscape view. The energy landscape view likens the energy landscape of a protein to a funnel, with the native structure at its global minimum, and each molecule may follow a different microscopic route from the top to the bottom (energy landscapes are discussed in more detail later).

The many models of protein folding are not mutually exclusive; they try to grasp different aspects of folding, and experimental results give some support to each model we mentioned. A newer model, named "nucleation–condensation model," is an attempt to unite the features of both the framework and the hydrophobic collapse mechanisms (Fersht, 1995; Fersht, 1997). In this model, long-range and other native hydrophobic interactions form in the transition state to stabilize the otherwise weak secondary structure. The framework and the hydrophobic collapse models are viewed as two extremes of the nucleation–condensation mechanism; most proteins fold by a mechanism that is somewhere between the two extremes, i.e., secondary structure and hydrophobic interactions form nearly simultaneously and synergistically (Daggett and Fersht, 2003).

3 Folding Mechanisms and Kinetics

A unified view of protein folding should be general enough to interpret the diverse experimental findings of the field. Thermodynamics offers such a universal approach. Thermodynamic systems in equilibrium occupy the states with lowest Gibbs free energy at constant pressure and temperature. The Gibbs free energy (G) consists of an enthalpy and an entropic term

$$G(q) = H(q) - T(q),$$

where H is the enthalpy, T the absolute temperature, and S the entropy of the protein, and q represents the reaction coordinate used to describe the progress of the protein advancing from the unfolded toward the native state. Under physiological conditions, proteins maintain their native structure because the favorable enthalpic term arising from the solvent and protein interactions exceeds in magnitude the unfavorable entropic term, and therefore the native state has a smaller Gibbs free energy than the denatured state. The stability of the protein depends on the solvent–solvent, protein–solvent, and protein–protein interactions. These interactions depend on the intensive parameters that describe the thermodynamic state of the system.

The enthalpic and the entropic terms are large, but of opposite sign, and almost cancel each other. The Gibbs free-energy difference between the biologically active and denatured states of the proteins is rather small (Scharnagl et al., 2005). Proteins are stable only within a narrow range of conditions and can be denatured by changing virtually any of the intensive parameters (Shortle, 1996). Experiments prove that proteins can be unfolded by heat (Tsai et al., 2002; Prabhu and Sharp, 2005), cold (Franks, 1995; Kunugi and Tanaka, 2002), high pressure (Smeller, 2002; Meersman et al., 2006), extreme pH (Puett, 1973; Fitch et al., 2006), and addition of salts (Pfeil, 1981).

Studies of protein stability and folding systematically change one or more of the intensive parameters and follow the kinetics of the change and/or the shift of equilibrium. There is a broad selection of methods that can be used to follow the structural changes of the proteins, including fluorescence (Isenman et al., 1979; Vanhove et al., 1998), phosphorescence (Mersol et al., 1993; Mazhul' et al., 2003), circular dichroism (Kelly and Price, 2000), infrared spectroscopy (Fabian and Naumann, 2004; Ma and Gruebele, 2005),

nuclear magnetic resonance (Englander and Mayne, 1992; Kamatari et al., 2004), and mass spectroscopy (Miranker et al., 1996; Konermann and Simmons, 2003).

Both theoretical and experimental results indicate that a single reaction coordinate in general is not enough to describe protein folding, and multiple reaction coordinates must be used. (Becker and Karplus, 1997; Ma and Gruebele, 2005). Finding the adequate reaction coordinates for protein folding is not straightforward. Several kinetic and thermodynamic coordinates have been used to describe the "nativeness" of a given protein state. Thermodynamic reaction coordinates use a thermodynamic parameter, e.g., Gibbs free energy and/or entropy, to define the distance between the native state and the actual state of a protein. The kinetic reaction coordinate measures the time needed for the protein to reach the native state from a given starting state. An important thermodynamic reaction coordinate often used to describe the folding process is the number of native contacts present in the conformation, which proved useful in interpreting simple folding processes. Thermodynamic reaction coordinates, however, are inadequate to describe folding dominated by kinetic traps because they completely ignore the Gibbs free-energy barriers separating the different states (Sali et al., 1994; Wolynes et al., 1995; Chan and Dill, 1998).

The Gibbs free-energy barrier to folding is determined by the unfavorable loss in configurational entropy upon folding and the gain in stabilizing native interactions. Starting from the unfolded protein, the polypeptide chain has to fold partially in order to bring together the residues that need to form the contacts stabilizing the native structure. The constrained polypeptide chain has smaller entropy, which means higher Gibbs free energy. As native contacts form, the enthalpy term decreases, the protein is stabilized. The rate-limiting step in the folding process is the formation of the transition state, i.e., the conformation that has the highest Gibbs free energy on the folding pathway (Chan and Dill, 1998; Lindorff-Larsen et al., 2005).

The simplest model for unfolding and refolding involves a single cooperative folding step, in which the unfolded (U) and folded (F) states of the protein interconvert: $U \leftrightarrow F$. This simple mechanism well describes the folding of several small proteins (Gillespie and Plaxco, 2004). The formation of a contact between two residues in the transition state involves an entropic cost which depends on the sequence separation of the two residues: the longer the chain between them the greater the entropic cost, and this entropic cost contributes to the height of the Gibbs free-energy barrier between the unfolded and the folded state. If nonnative interactions play a marginal role in the transition state, it is possible to estimate the folding kinetics from the average sequence separation of the contacts in the native structure (Plaxco et al., 1998; Grantcharova et al., 2001; Zarrine-Afsar et al., 2005).

Intermediate structures were observed to accumulate during the folding of many proteins (Englander, 2000). Such intermediate states are trapped structures that have low Gibbs free energy. Mass action models that involve one or more intermediate states were constructed to explain more complex folding kinetics. Mass action models distinguish between "on-pathway" and "off-pathway" intermediates depending on whether the intermediate is on the folding pathway between the unfolded and native states (Baldwin, 1996). Off-pathway intermediates often correspond to misfolded structures that must completely or partially unfold to allow formation of the native fold (Evans et al., 2005). The most general theory of protein folding is a statistical mechanical model that uses the concept of the energy landscape, which is discussed in more detail later. Here we only want to clarify that in the energy landscape view, there is no clear distinction between on- and off-pathway intermediates. Folding mechanisms involving these two types of intermediates only differ in the distribution of traps on a Gibbs free-energy landscape (Onuchic and Wolynes, 2004; Jahn and Radford, 2005).

AU1

Energy landscape theory of protein folding predicts that the enthalpic and the entropic term of the transition-state Gibbs free energy can cancel each other, leading to folding that lacks an activation barrier (Bryngelson et al., 1995). Although such downhill folding has indeed been found experimentally, it seems to be atypical, probably because such proteins are evolutionary unfavorable (Yang and Gruebele, 2004a). The probable reason for this is that proteins that fold downhill lack the barrier that prevents partial unfolding, and thus become more prone to aggregation (Yang and Gruebele, 2003; Gruebele, 2005).

The folding process is usually not restricted to a narrow path in conformational space. Each molecule may follow a different path, and the molecule population can be very heterogeneous. Mass action models are inadequate to describe these folding processes, and the more complex energy landscape view must be used.

Heterogeneous folding ensembles give rise to stretched folding kinetics and/or probe-dependent observation of the kinetics (Sabelko et al., 1999; Osvath et al., 2003; Ma and Gruebele, 2005; Osvath et al., 2006).

4 Molten Globules and Other Compact Denatured States

As mentioned previously, Levinthal (1969) postulated the existence of folding pathways in 1968. In an effort to find the specific intermediates along the folding pathway, folding studies were performed on several globular proteins (Wong and Tanford, 1973; Kuwajima et al., 1976; Robson and Pain, 1976; Nozaka et al., 1978). Equilibrium intermediates were reported but it was not clear whether these are related to the specific folding intermediates. However, the equilibrium intermediates characterized in different proteins were found to be remarkably similar to each other: all of them had native-like secondary structure and were compact, but lacked a specific tertiary structure. In 1983, Ohgushi and Wada (Ohgushi and Wada, 1983) proposed that the equilibrium intermediates belong to a common physical state of globular proteins, and they termed this the "molten globule" state. After Kuwajima and coworkers (1985) and Ikeguchi and coworkers (1986) had shown that the molten globule state of α-lactalbumin is identical to its transient folding intermediate, researchers started to study molten globules with renewed interest. Molten globule states could be generated using mild denaturing conditions (low or high pH, moderate concentrations of denaturants, high temperature, and various salts) in about 20–25 different proteins such as α-lactalbumin, carbonic anhydrase B, β-lactamase, ribonuclease A, T4 lysozyme, cytochrome *c*, apomyoglobin, and staphylococcal nuclease (Ptitsyn, 1995).

4.1 The Structure of the Molten Globule

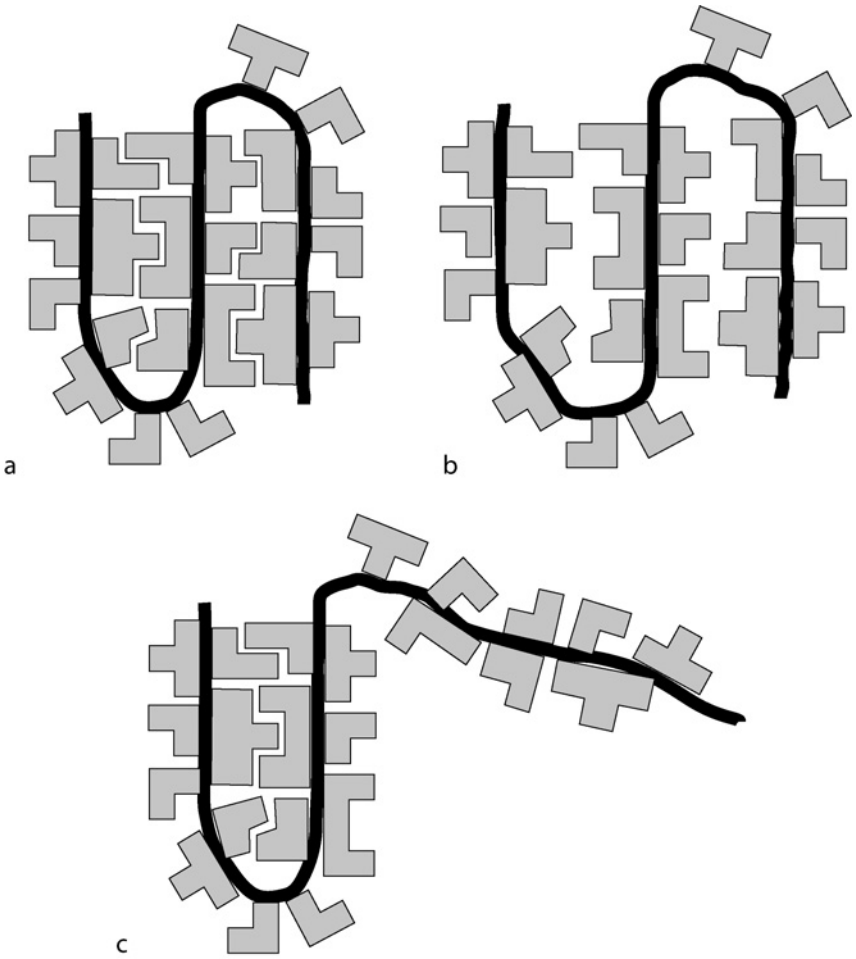
The common characteristics of the molten globule state as described by Kuwajima (Arai and Kuwajima, 2000) are (1) the presence of a significant amount of secondary structure, (2) the absence of most of the specific tertiary structure produced by the tight packing of side chains, (3) compactness of the protein molecule with a radius of gyration 10%–30% larger than that of the native state, and (4) the presence of a loosely packed hydrophobic core that increases the hydrophobic surface accessible to solvent. The experimental techniques typically used to detect molten globules are (1) far- and near-UV CD spectra that detect the secondary and tertiary structures of a protein, (2) hydrodynamic methods such as viscosity measurement and molecular sieve chromatography that determine the molecular size of the protein, and (3) hydrophobic dye (typically ANS, 8-anilino-naphtalene-1-sulfonate) binding experiments that detect the formation of a loose hydrophobic core and estimate the extent of hydrophobic area exposed to the solvent.

These techniques, however, only provide information about the average structural properties of the protein molecule. More advanced techniques that provide a more detailed picture about the molten globule were not available until the 1990s. Therefore, the exact nature of molten globule structure had been a matter of some debate. ❯ *Figure 10-2* shows an illustration of two different models for the molten globule. The traditional view of the molten globule (Shakhnovich and Finkelstein, 1989; Ptitsyn, 1992) assumes that the backbone of the polypeptide chain essentially has a (fluctuating) native-like fold and the disorder is in the side chains. However, this view was criticized on the basis of thermodynamic arguments and energy landscape theory (Dill et al., 1995; Privalov, 1996). Privalov (1996) argued that molten globule states are either misfolded structures or states where one portion of the structure is already folded and another one is still unfolded. Dill and coworkers (1995) stated on the basis of simulations of simplified models that backbone and side-chain degrees of freedom are strongly coupled, therefore a state where the backbone is ordered and the side chains are disordered is unlikely.

The new experimental techniques developed in the 1990s, such as stopped-flow circular dichroism, pulsed hydrogen exchange, X-ray scattering, and mutational approaches, have allowed characterization of the molten globule states in great detail (Hughson et al., 1990; Dobson, 1994; Carlsson and Jonsson, 1995; Carra and Privalov, 1996; Dabora et al., 1996; Dyson and Wright, 1996; Kataoka and Goto, 1996; Kuwajima, 1996; Song et al., 1998; Chakraborty et al., 2001; Demarest et al., 2001; Ramboarina and

Figure 10-2

Illustration of different models for the structure of the molten globule. The backbone of the protein is represented by a *thick line* and side chains are shown as pieces of *various shapes* hanging from the backbone. (a) The native state of the protein, with all the buried side chains fitting closely together like the pieces of a jigsaw puzzle. (b) The side chain molten globule model: the fold of the backbone is native-like but the side chains are only loosely packed. (c) One-half of the molecule is fully folded, with the specific side chain interactions present, while the other half is completely unfolded



Redfield, 2003; Redfield, 2004). These studies confirmed the view that the molten globule state is indeed more heterogeneous than previously thought: one part of the structure is more organized and native-like while other portions are less organized. However, there is a great deal of variety among different proteins in the exact structure of the molten globule state.

4.2 The Role of the Molten Globule in Folding

The study of molten globules was strongly motivated by the idea that the molten globule state could be identical to the transient intermediate during folding. This notion was supported by kinetic circular dichroism measurements on α-lactalbumin (Ikeguchi et al., 1986). However, later theoretical studies

suggested that the experimentally observed folding intermediates may not be productive but kinetically trapped, misfolded species (Dill and Chan, 1997). Careful kinetic measurements during the refolding of several proteins including interleukin-1β, staphylococcal nuclease, and apomyoglobin (Heidary et al., 1997; Walkenhorst et al., 1997; Jamin and Baldwin, 1998; Maki et al., 1999) have provided firm evidence that, contrary to the theoretical predictions, the molten globule state is the productive on-pathway folding intermediate in most cases. An attempt to resolve the apparent contradiction between theory and experiment introduced the hierarchical folding model (Arai and Kuwajima, 2000) where the folding of a protein occurs in two stages: (1) formation of the molten globule state from the fully folded state and (2) formation of the native state from the molten globule state.

5 Simulations of Protein Folding and Unfolding

Computational models and simulations have greatly advanced our understanding of protein folding. The information from such studies is complementary to experiments. In fact, there is a synergy between theory and experiment: theory provides testable models and experiments provide the means to test and validate the models. The outcome from this combination is a much richer view of the system in question than what either approach could provide alone. In particular, simulations can help identify or predict transition and intermediate states along the folding pathway, provide predictions of the rate of folding and in some cases, predict the final, folded structure.

Simulating protein folding presents a significant challenge. Small proteins typically fold in the several microseconds to seconds timescale; detailed atomistic simulations, however, are currently limited to the nanosecond to microsecond regime. Therefore, simulation of folding requires either simplified models or special sampling methods, both of which introduce new approximations.

5.1 All-Atom Models

The most straightforward approach to simulating protein folding and unfolding is to use an all-atom model with a force field like AMBER or CHARMM and apply traditional molecular dynamics simulation. These force fields describe the energies of the deformations of covalent bonds as well as van der Waals interactions, charge–charge interactions, hydrogen bonds, and so on. Traditional molecular dynamics numerically solves Newton's equations of motion by calculating the forces acting on atoms and computing accelerations, velocities, and atomic displacements. Temperature is assigned to the system by assigning appropriate velocities to the atoms. To simulate unfolding, the simplest method of increasing sampling is to increase the temperature of the simulation to 498 K or more. At these temperatures, the native structure of the protein is usually lost within a few nanoseconds. This technique has been applied to numerous examples, including bovine pancreatic trypsin inhibitor (Kazmirski and Daggett, 1998a), lysozyme (Kazmirski and Daggett, 1998b), myoglobin (Tirado-Rives and Jorgensen, 1993), barnase (Wong et al., 2000), ubiquitin (Alonso and Daggett, 1998), the SH3 domain (Tsai et al., 1999), etc. Features of the unfolding process such as the transition-state ensemble or the unfolded ensemble have shown remarkable agreement with experimental results (Day and Daggett, 2003).

AU2

A direct simulation of folding is a much harder problem. The longest continuous all-atom molecular dynamics simulation so far is still the 1 μs simulation of villin headpiece subdomain performed by Duan and Kollman in 1998 (Duan and Kollman, 1998). In this simulation, the chain sampled a large number of conformations after initial collapse, and near-native structures appeared but the true native conformation was not reached. Although processor speeds have dramatically increased since 1998, computational power is still insufficient to allow a meaningful all-atom simulation of the entire folding process. Even if the native state could be reached in a single trajectory, multiple simulations would have to be performed to construct a believable folding pathway.

An interesting alternative is the massively distributed method employed in the "Folding at Home" project (Pande et al., 2003). A large number of computers run the simulation in parallel. As soon as a

transition is detected (as a momentary surge in the heat capacity) in one of the simulations, all computers receive a copy of the posttransition conformation and the simulation is continued until the native conformation is reached. Although this approach has been criticized as being flawed (Fersht and Daggett, 2002), it has been successfully applied to fold several small, fast-folding proteins (Snow et al., 2004; Sorin and Pande, 2005).

5.2 Simplified Models

In simplified (coarse-grained) models (Dokholyan, 2006), effective particles (beads) represent amino acids or groups of atoms. An empirical potential function, usually derived from protein structures, is used to describe the interaction between these beads. The shape of this potential is often very simple, such as a square-well function. In many simplified models, the positions of the beads are restricted to points on a lattice. Perhaps the most minimal model is the one where there are only two types of amino acids: hydrophobic and polar, and the chain is restricted to a two-dimensional lattice (Dill et al., 1995).

Smaller on-lattice model proteins allow an exhaustive enumeration of all possible states of the given system. This approach allows a complete thermodynamic description of the phase space and has greatly enhanced our understanding of protein folding. The funnel view and the concept of energy landscapes (see *Sect. 10-6*) arose directly from the exhaustive sampling allowed by these minimal models (Bryngelson et al., 1995).

In the case of larger, more complex simplified models, exhaustive enumeration of states is not possible. Monte Carlo is a common choice for simulating simplified models. In Monte Carlo simulation, small moves are generated randomly and accepted or rejected based on the energy of the new conformation. This is often performed in the framework of advanced sampling schemes such as Replica Exchange Monte Carlo, where several replicas of the system are simulated at various temperatures (Kihara et al., 2001; Pokarowski et al., 2003). A more recent simulation approach, termed discrete (or discontinuous) molecular dynamics (DMD) (Smith and Hall, 2001; Ding and Dokholyan, 2005), extends the accessible simulation time by using long integration time steps with approximated energy functions. Simple models like this start showing remarkable success. Recently, Trp-cage, a 20-residue miniprotein has been folded to a conformation very close to the experimental structure (Ding et al., 2005a). It is believed that this technique will be applicable to larger proteins.

5.3 Multiscale Modeling

Approaches using simplified, coarse-grained models can be combined with fine-grained, all-atom simulations in what is called multiscale molecular modeling. Bradley and coworkers (2005) reported high-resolution structure prediction for proteins up to 85 residues using a multiscale approach that sampled low- and high-resolution conformations. DMD simulations combined with all-atom, traditional molecular dynamics have been used to simulate the formation of a β helix (Khare et al., 2005) and to identify the transition state of the SH3 domain (Ding et al., 2005b). Although it remains to be seen whether the force fields used are transferable to larger proteins, multiscale modeling has the potential to break the 1 μs barrier of direct folding simulations.

6 Free-Energy Landscapes of Proteins

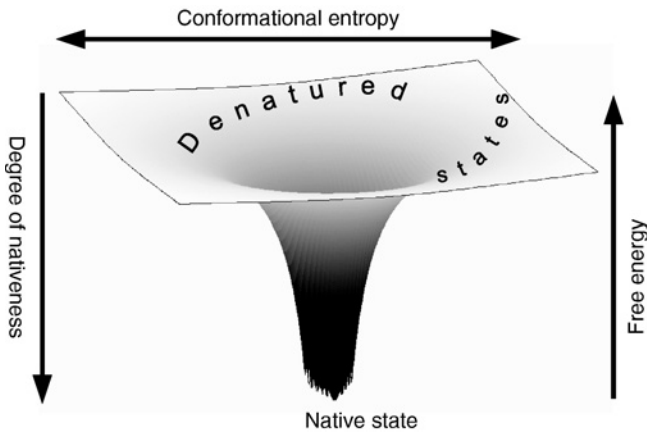
Energy landscapes are mathematical devices that help us understand the microscopic behavior of a molecular system (Bryngelson et al., 1995). An energy landscape of a system with n degrees of freedom is an energy function $F(\mathrm{x}) = F(x_1, x_2, \ldots, x_n)$ where $x_1, \ldots, x_n$ are variables specifying the microscopic state of the system (Dill, 1999). In the case of a protein, $x_1, x_2, \ldots, x_n$ can be, for instance, all the dihedral angles of the chain, thus specifying a single conformation of the protein. $F(\mathrm{x})$ then is usually defined as the free

energy of the protein in the given conformation, where the entropic part of the free energy comes from all possible solvent configurations. Thus, $F(x)$ is the free energy of a microstate, not a macrostate, because it does not include the chain conformational entropy. The stable conformation of the protein can be found by determining the set of values $x_1, x_2, \ldots, x_n$ (i.e., the conformation) that gives the minimum value of the free-energy function.

Although energy landscapes are, by definition, high-dimensional surfaces, they are often pictured as a surface in three dimensions. In these pictures, the vertical axis represents the free energy and the horizontal axes represent the conformational degrees of freedom of the polypeptide chain. Random heteropolymers, such as a random sequence of amino acids, have a very rugged energy landscape with many local minima (Plotkin et al., 1996). Systems like this easily get trapped in one of the local minima and usually do not have a well-defined, single, stable conformation. Real proteins, however, are not random sequences; evolution has optimized their sequences so that they quickly and efficiently fold into a well-defined three-dimensional conformation (Onuchic and Wolynes, 2004). In a real protein, most of the interactions that can form between parts of the chain are mutually supportive and cooperatively lead to a low-energy structure which is therefore "minimally frustrated." This "principle of minimal frustration" (Bryngelson and Wolynes, 1987), gleaned from simplified models of proteins and the theory of spin glasses, led to the realization that the energy landscape of a real protein should be shaped like a funnel (*Figure 10-3*).

Figure 10-3

Schematic representation of a funnel-shaped energy landscape. The width of the funnel represents the conformational freedom of the chain. The vertical axis represents the free energy; as free energy decreases, the nativeness of the chain increases. Denatured (unfolded) states are at the top of the funnel while the native state is the global minimum. There is some ruggedness in the energy landscape near the native state



A funnel-shaped energy landscape means that the free energy of a structure depends on how close it is to the native state: the closer it is to the native structure the lower its free energy. Also, the top of the funnel, representing the nonnative states, is wide (the conformational entropy is high), and it narrows as one gets closer to the bottom: near-native states represent more compact conformations, and therefore have low conformational entropy.

The fact that the energy landscape of a real protein is essentially funnel-shaped has several important consequences (Wolynes, 2005). First, the native structure should be robust against mutations (Nelson and Onuchic, 1998). A point mutation represents a small perturbation to the energy landscape; therefore, the basic shape of the funnel and the location of the global minimum cannot change much: the mutant protein will fold into essentially the same structure as the wild-type protein. This structural robustness underlies the most commonly used method of predicting structures: homology modeling (Zhang and Skolnick, 2004).

If sequence analysis shows two sequences to be evolutionarily related, one can be almost certain that the two structures are essentially the same.

Another important consequence of the funnel shape is that the folding rate and folding mechanism of a protein is largely determined by its native-state topology (Baker, 2000). Assuming that the energy landscape is a perfect funnel, its shape is completely determined by the topology of the native state, and this shape completely determines the folding mechanism. Of course, the energy landscape of a real protein is never a perfect funnel: it is always somewhat rugged, and the exact layout of the small heaps and valleys on the landscape will influence the details of the folding mechanism. Still, for many proteins, the landscape is close enough to a perfect funnel to allow the development of methods that can successfully predict various features of the folding mechanism from just the native structure as input (Alm and Baker, 1999; Galzitskaya and Finkelstein, 1999; Munoz and Eaton, 1999). The folding rate can be predicted from just the contact order (the average sequence separation between residues that make contacts in the native structure). Using a perfectly smooth, funnel-shaped energy function based on the Go model (Go, 1983) (where only native contacts contribute to the free energy), transition-state structures, folding nuclei, and Φ values have been predicted for several small proteins such as CheY, CI-2, barnase (Galzitskaya and Finkelstein, 1999), λ-repressor, and SH3 domain (Alm and Baker, 1999), with surprisingly good agreement with experiments.

The concept of energy landscapes also helps us understand some of the more complex protein-folding scenarios. The folding of many proteins involves slow steps, bottlenecks, and multiple, kinetically distinguishable stages (Wolynes et al., 1995). These phenomena can be rationalized by noticing that energy landscapes are often not perfect, smooth funnels but are rugged and bumpy (Dill, 1999). Slow steps in the folding process can arise from climbing an uphill slope after being trapped in local minima. This scenario may appropriately describe the folding of β-lactoglobulin, a predominantly β-sheet protein, which passes through a helical phase as it folds (Hamada et al., 1996). Because the landscape view is a microscopic view of the folding process, each molecule may follow a different route on the energy landscape and may encounter different obstacles in its way. For example, in the folding of hen egg white lysozyme, a subpopulation of the molecules folds fast while another subpopulation folds quite slowly (Radford et al., 1992). In the landscape view, it is easy to see how molecules starting from one region of the top of the funnel may "ski down" unhindered while molecules starting from another region may get trapped behind a mountain range. Bottlenecks do not always involve uphill climbing though, they can be entropic barriers too; in this case, the aimless search for a downhill route on a large, level field will limit the rate of folding (Dill and Chan, 1997). This example shows that bottlenecks are typically ensembles of widely different conformations rather than well-defined, single conformations (Chan and Dill, 1998). The landscape view of protein folding also gives us a clue to chaperone function: to get a misfolded protein to fold correctly, no specific recognition is needed; it is sufficient just to move it to the top of the funnel where it can restart the downhill search for the folded conformation (Chan and Dill, 1996; Todd et al., 1996).

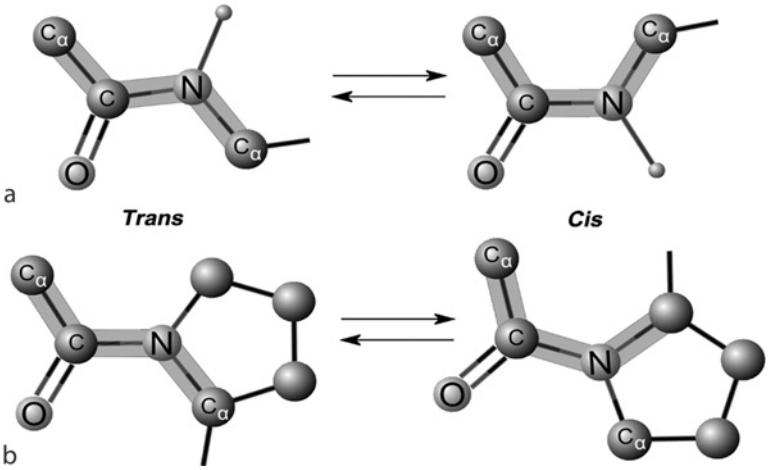
7 Traps on the Folding Pathway

The folding protein faces several obstacles in translating the information encoded in the amino acid sequence into the three-dimensional construct of the biologically active structure. The *cis–trans* isomerization of peptide bonds and the formation of disulfide bridges are slow steps that can form bottlenecks in the protein-folding reaction (Balbach and Schmid, 2000; Creighton, 2000). Under certain circumstances, the two processes can be linked and the formation of the correct disulfide bonds is facilitated in the presence of peptidyl-prolyl *cis–trans* isomerase (Schonbrunner and Schmid, 1992).

7.1 Backbone Isomerization

The peptide bond between nonproline amino acids is much more stable in the *trans* than in the *cis* conformation (❯ *Figure 10-4a*). The difference in stability arises from interactions between the C_i^{α} and H_i^{α} atoms with the C_{i+1}^{α} and H_{i+1}^{α} atoms, from electrostatic interactions between the O_i and C_{i+1} atoms

Figure 10-4
(a) The *cis* and *trans* states of nonprolyl peptide bonds. (b) The *cis* and *trans* states of prolyl peptide bonds



and from conformational entropy (Zimmerman and Scheraga, 1976; Wedemeyer et al., 2002). The large difference in the stability of the two isomers keeps the *trans* form 100 to 1000 times more populated than the *cis* isomer (Ramachandran and Mitra, 1976; Jorgensen and Gao, 1988; Scherer et al., 1998). Because of the high number of peptide bonds in the protein, the fraction of protein molecules that have *cis* peptide bonds in the denatured ensemble can be significant. Since isomerization of peptide bonds is a slow process, nonnative isomers of peptide bonds can cause slow phases in protein folding. It has been shown that *cis–trans* isomerization of nonprolyl peptide bonds can give rise to significant slow folding phases (Eyles, 2001). Nonprolyl *cis* peptide bonds are energetically unfavorable and are very rare in the native structures of proteins (Stewart et al., 1990). Isomerization of the *trans* peptide bond accumulated in the denatured state into the native *cis* conformation of these proteins can become the rate-limiting step of the folding reaction (Odefey et al., 1995).

Proline residues are a much more common source of kinetic complications during folding. The X-Pro peptide bond (where X can be any amino acid) is the only peptide bond for which the stability of the *cis* and *trans* conformations is comparable. The *cis–trans* isomerization of X-Pro peptide bonds is a widely recognized rate-limiting factor, which can induce additional slow phases in protein folding (Brandts et al., 1975; Wedemeyer et al., 2002).

The otherwise strong preference of the peptide bond to be in *trans* state does not apply to the X-Pro bond, mostly because of the steric symmetry between the C^{α} and the C^{γ} atoms of proline (*Figure 10-4b*). The entropy change accompanying the *cis–trans* transition and the electrostatic interactions between the O_i and C_{i+1} atoms are also different in the X-Pro bonds than in other peptide bonds (Zimmerman and Scheraga, 1976; Wedemeyer et al., 2002). As a result, the *trans* conformation of X-Pro bonds is only marginally more stable than the *cis* conformation in unfolded polypeptides. In the unfolded protein, a mixture containing both conformations is present, with 10%–30% of the X-Pro bonds in the *cis* state (Cheng and Bovey, 1977; Grathwohl and Wuthrich, 1981).

In the native conformation of a protein, intraprotein interactions will stabilize either the *cis* or the *trans* isomer for most of the X-Pro bonds. The molecules that contain incorrect isomers in the unfolded ensemble must undergo isomerization during the folding reaction (Kiefhaber et al., 1990a, b; Texter et al., 1992). Consequently, the presence of proline residues leads to a number of proline isomerization reactions that must occur before folding can complete. Both experimental and theoretical findings show that there is a high energy barrier for isomerization (16–20 kcal/mol in model compounds) of X-Pro bonds (Balbach and Schmid, 2000; Kang and Choi, 2004). This results in characteristic times of 10–1000 s for the

conformational flipping at room temperature (Brandts et al., 1975). Mutagenesis studies have indeed shown that specific proline residues can often be assigned to slow recovery phases from misfolded states (Evans et al., 1987; Kelley and Richards, 1987; Wood et al., 1988; Herning et al., 1991; Kiefhaber et al., 1992; Wu and Matthews, 2002; Street et al., 2005).

The high energy barrier for isomerization reflects a partial double-bond character of the X-Pro bond (Balbach and Schmid, 2000). In vivo, enzymes (e.g., peptidyl-prolyl *cis–trans* isomerases) facilitate folding to the native structure by lowering the energy barrier for isomerization (Schmid et al., 1993; Stein, 1993; Gothel and Marahiel, 1999; Shaw, 2002).

Laser-induced temperature-jump measurements indicate that the role of proline residues in protein folding is more complex. Prolines can have opposite effects on the slow and fast steps of the folding kinetics. The presence of proline residues leads not only to additional slow phases, but also modifies the millisecond and sub-millisecond dynamics of the protein. The X-Pro bonds do not isomerize on the millisecond timescale. This increased backbone rigidity can speed up the fast folding steps of the ensembles that contain the proline in the native-like isomerization state (Osvath and Gruebele, 2003).

7.2 Formation of Disulfide Bridges

Formation of the correct native-like disulfide bridges can also hinder protein folding. Disulfide bridges between pairs of cysteines are part of the native structure for many proteins. The formation of disulfide bridges is a prerequisite of the proper folding and biological function in these proteins. Disulfide bonds increase the thermodynamic stability of the native structure by establishing conformational constraints within the protein (Creighton, 2000).

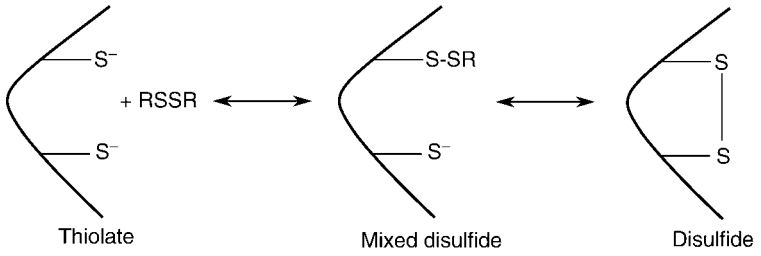
It has been shown that several proteins start to fold during synthesis and disulfide bond formation begins in the emerging chain (Bergman and Kuehl, 1979; Peters and Davidson, 1982; Braakman et al., 1991; Braakman et al., 1992). Folding and disulfide bridge formation is completed after the end of the translation (Wedemeyer et al., 2002).

Depending on the number of the cysteine residues in a polypeptide sequence and on the native structure, misfolding can occur owing to errors in disulfide pairing. Disulfides formed randomly in early folding intermediates may cause kinetic complications during the later folding steps (Creighton, 1979; Konishi et al., 1982). Repairing of the errors is part of the folding process and occurs in a trial and error manner (Chatrenet and Chang, 1992; Schwaller et al., 2003). Since disulfide bonds constrain the free movement of the polypeptide chain, the formation of native disulfide pairs can also prevent correct folding. It has been shown that native-like disulfide bridges formed too early during folding must be broken up to allow the folding process to continue, and they can reform at a later stage (Creighton and Goldenberg, 1984).

Formation of disulfide bonds requires the presence of an oxidizing agent or a disulfide reagent such as glutathione or dithiothreitol. When a disulfide reagent is present, disulfide bond formation occurs by a thiol–disulfide exchange reaction. This is a two-electron redox reaction in which a disulfide reagent takes one electron from each cysteine thiolate, and a disulfide bond is formed (❯ *Figure 10-5*) (Wedemeyer et al., 2000). The first step is the result of a nucleophilic attack by a cysteine thiolate on the disulfide reagent. The rate of this reaction is determined by the reactivity of the cysteine thiolate, the presence and nature of the disulfide reagent, pH, temperature, ionic strength, and cosolvents. In the second step, the mixed disulfide is broken up by another nucleophilic attack by the second cysteine thiolate. As a result, the external thiol is replaced by a protein thiolate and an intraprotein disulfide bond is formed. The prerequisite of the second step is a conformational change in the protein that brings the two cysteine residues together, thus the redox reaction is coupled to the folding of the protein (Creighton, 1997; Bulaj, 2005).

The disulfide pairing during in vivo folding is helped by protein disulfide isomerases (Bulleid and Freedman, 1988; Freedman, 1989; Ellgaard and Ruddock, 2005). These enzymes correct the disulfide pairing mistakes by catalyzing the reshuffling of disulfide bridges so that the native pairing can emerge (Gilbert, 1997). The process leads to a biologically active structure, which is resistant to further rearrangement (Walker and Gilbert, 1997).

Figure 10-5
Scheme of the redox reaction that leads to the formation of disulfide bonds in proteins



8 Transition States on the Folding Pathway

To gain insight into the folding mechanism, it is essential to characterize the energy landscape of the protein. This includes the determination of the structures and energies of the protein in the traps and the transition states of the landscape (Gruebele, 2002). The local minima of the energy landscape act as traps that are responsible for the buildup of kinetic intermediate states. Since the protein accumulates in measurable quantities in the intermediate states, the intermediate structures can be experimentally studied (Roder et al., 2000). Transition states are usually not populated in detectable quantities. The only way to gain information about the structural and energetic properties of the transition states is through kinetic studies of the folding reaction (Daggett and Fersht, 2000).

The kinetics of the transition between different states is determined by the escape rates from local minima of the landscape. In simple cases, protein folding can be described as a diffusive process over a barrier determined by the energy landscape. For barriers that are larger than the thermal energy $k_B T$, the folding rate predicted by transition-state theory can be calculated by a Kramer-like equation (Onuchic et al., 1996)

$$k = v \exp\left(-\Delta G^{\dagger} \Big/ k_B T\right).$$

Here k denotes the rate of the conformational transition, v is a coefficient that depends on a number of things including the shape of the barrier and solvent viscosity, $\Delta G^{\dagger}$ is the Gibbs free-energy height of the transition state, k_B is Boltzmann's constant, and T is the absolute temperature. Transition-state theory thus allows us to determine the Gibbs free energy of the transition state using simple kinetic measurements. Gaining information about the structure of the transition state is a more intricate problem.

To date, the only way to learn about the structure of the transition state is Φ-value analysis, a method that uses site-directed mutagenesis to map out the residue–residue contacts present in the transition state (Matthews and Hurle, 1987; Fersht et al., 1992). The calculated Φ-value compares the change in the folding rate and the change in the stability of the protein caused by a specific mutation (Clementi et al., 2000)

$$\Phi = \left(-RT \ln(k_{\mathrm{mut}}/k_{\mathrm{wt}})\right) \Big/ \left(\Delta G^{\circ}_{\mathrm{mut}} - \Delta G^{\circ}_{\mathrm{wt}}\right).$$

Here R is the gas constant, T is the absolute temperature, and k_{mut} and k_{wt} are the folding rates of the mutant and wild-type protein, respectively. $\Delta G^{\circ}_{\mathrm{mut}}$ and $\Delta G^{\circ}_{\mathrm{wt}}$ represent the stabilities of the mutant and of the wild-type protein, i.e., the Gibbs free-energy difference between the folded and the unfolded state. If the unfolded state is assumed to be a randomly fluctuating chain, its free energy can be taken as the solvation free energy. However, it has been shown that denatured proteins can have some residual structure (Smith et al., 1996; McCarney et al., 2005), and a determination of their free energies becomes more complicated. For the purposes of Φ-value analysis, however, we can just use the Gibbs free energy of the unfolded state as a reference, i.e., define it as zero.

The expression for Φ can be simplified if the following two conditions are met:

(1) The folding rate can be calculated from the activation energy using a Kramer-like exponential dependence, thus folding can be treated as a simple cross-barrier diffusive reaction and (2) The folding

mechanism is not altered significantly by the mutation, thus the parameter v reflecting the barrier shape and the configurational diffusion of the protein is insensitive to the mutation.

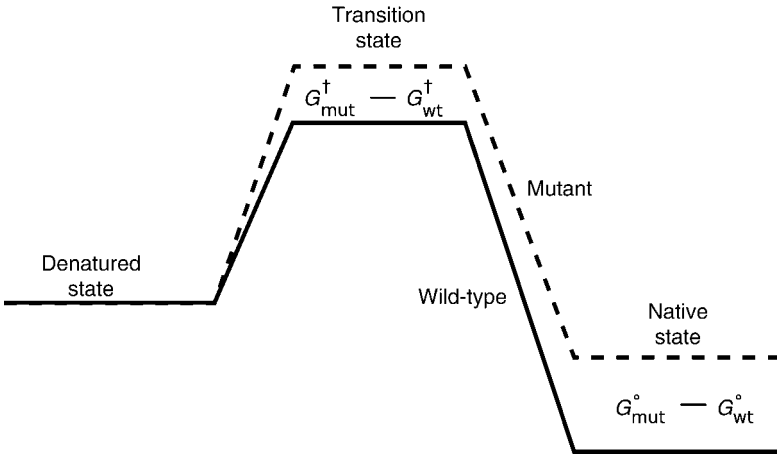
If the above assumptions hold, the new expression for Φ is:

$$\Phi = \left(\Delta G^{\dagger}_{\mathrm{mut}} - \Delta G^{\dagger}_{\mathrm{wt}}\right) \Big/ \left(\Delta G^{\circ}_{\mathrm{mut}} - \Delta G^{\circ}_{\mathrm{wt}}\right),$$

where $\Delta G^{\dagger}_{\mathrm{mut}}$ and $\Delta G^{\dagger}_{\mathrm{wt}}$ denote the activation Gibbs free energies for the mutant and the wild-type protein, respectively. This way, the Φ-value compares the Gibbs free-energy change introduced by the mutation in the transition state and in the native state (*Figure 10-6*).

Figure 10-6

Schematic representation of the Gibbs free energies important in the Φ-value analysis



In order to be able to extract structural information from Φ-value analysis, two more conditions have to be met (Fersht et al., 1992; Clementi et al., 2000):

1. The folding pathway is not altered significantly by the mutation, thus the intermediate and transition states are the same for the mutant and the wild-type protein.
2. The folded region of the transition state has a native-like structure.

The above conditions were found to be true for several proteins, and Φ-value analysis was used to characterize transition states and short-lived intermediates of several folding pathways (Matouschek et al., 1990; Serrano et al., 1992; Grantcharova et al., 1998; Villegas et al., 1998; Fulton et al., 1999; Goldenberg, 1999; Raschke et al., 1999; Crane et al., 2000; Bulaj and Goldenberg, 2001; Capaldi et al., 2002; Paci et al., 2003; Hubner et al., 2004; Lindorff-Larsen et al., 2004; Hubner et al., 2005).

Usually, a detailed analysis of the mutation is necessary to identify the contacts disrupted in the mutant structure. The use of multiple mutations of the same residue can increase the accuracy of the prediction of the transition-state structure (Matouschek et al., 1995). Φ-values are normally between 0 and 1, but negative Φ-values corresponding to mutations that speed up the folding kinetics have also been observed (Yang and Gruebele, 2003; Yang and Gruebele, 2004b). The Φ-value is a valuable tool in detecting native contacts in transition-state structures, but in general, it is not proportional to the extent or strength of the native contact appearing in the transition state. This means that structural information can unambiguously be assigned only to $\Phi = 0$ and $\Phi = 1$ (Fersht and Sato, 2004).

A Φ-value close to 1 indicates that the free-energy change introduced by the mutation is almost identical for the transition state and the native state. This implies that the mutated residue already forms its native

contacts in the transition state. A Φ-value close to 0 indicates that the height of the Gibbs free-energy barrier of the transition state was not altered by the mutation. In this case, the mutated residue does not form native contacts in the transition state, and the environment of the mutated residue is probably denatured-like (Fersht and Sato, 2004).

It has been shown that for deletions of small hydrophobic residues, the Φ-value is roughly proportional to the extent of the native contact formation. Therefore, these residues are preferred as targets for mutations in Φ-value analysis (Fersht and Sato, 2004).

9 Folding of Multidomain and Multi-Subunit Proteins

9.1 Multidomain Proteins

A domain is a part of the polypeptide chain that forms a compact globular substructure with more interactions within itself than with other parts of the chain (Janin and Wodak, 1983). Most proteins longer than about 200 to 250 residues consist of several domains. When investigating the role of domains in folding, the first question to answer is whether a domain can fold by itself. Isolated domains can be produced by limited proteolysis or genetic engineering, and their folding can be studied by the same experimental methods as used for the whole proteins. Autonomous folding was demonstrated for the domains of several multidomain proteins including tryptophan synthase, β-lactamase, aspartokinase–homoserine dehydrogenase, plasminogen, phosphoglycerate kinase (Jaenicke, 1987), and several other domains (Sharma et al., 1990; Herold et al., 1991; Shoelson et al., 1993; Williams and Shoelson, 1993; Jecht et al., 1994). In many cases, it was shown that the stability of the domains in isolation is close to the stability when the other domains are also present (Garel and Dautry-Varsat, 1980; Muller and Garel, 1984; Novokhatny et al., 1984; Jaenicke, 1987; Tsunenaga et al., 1987; Rudolph et al., 1990; Missiakas et al., 1992). Interestingly, the rate of folding of isolated domains was found to be greater or the same as that of the same domains integrated within the intact protein (Teale and Benjamin, 1977; Dautry-Varsat and Garel, 1981; Blond and Goldberg, 1986; Tsunenaga et al., 1987; Missiakas et al., 1992). This suggests that the presence of the rest of the chain slows down the folding of a domain, e.g., by forming unfavorable interactions with it.

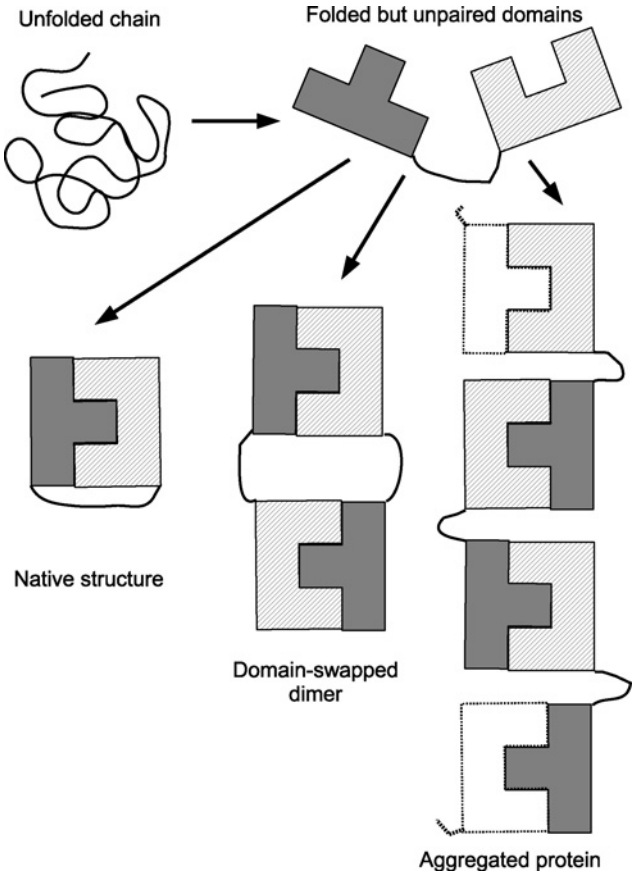
The folding kinetics of multidomain proteins is usually complex, showing an initial rapid phase characterized by large changes in several physical parameters (fluorescence, UV absorption, circular dichroism, etc.), followed by a second, slower phase by much smaller changes in the physical parameters (Jaenicke, 1987; Jaenicke, 1999). The intermediate that accumulates after the initial rapid phase appears largely folded but lacks some properties of the native structure: it is more labile to proteolysis, is not recognized by some antinative antibodies, and lacks catalytic activity. These properties only appear after the second, slower stage of folding. The findings suggest that the individual domains fold in the rapid phase of the folding process, and the second, slow step is the pairing or association of the already folded domains. This is also supported by the fact that in several cases the rate of folding was found to be inversely proportional to solvent viscosity, implying that movement of large, globular species is involved in the rate-limiting step of the folding process (Vaucheret et al., 1987; Chrunyk and Matthews, 1990). Also, in tryptophan synthase, single mutations in each domain were found to decrease the folding rate while the double mutant was able to fold at the same rate as the wild-type protein, suggesting that the rate-limiting step of folding involves the formation of the interface between the two domains (Tsuji et al., 1993). Thus, the folding process can be described by the following general scheme

$$\text{Unfolded} \xrightarrow[\text{folding}]{\text{Fast domain}} \text{Intermediate with folded but unpaired domains} \xrightarrow[\text{pairing}]{\text{Slow domain}} \text{Native}$$

The existence of an intermediate with folded but unpaired domains has important consequences. If the protein concentration is sufficiently high, the domain pairing step may occur between domains from two identical molecules, leading to the formation of a dimer instead of a monomer. The domains make the same interactions in the dimer as in the monomer, but the interactions are formed intermolecularly rather than intramolecularly. Higher-order oligomers may also form and the process may lead to fibril formation or

Figure 10-7

The folding of a hypothetical two-domain protein. The unfolded chain folds into two domains that are not yet paired. The fate of this intermediate depends on the protein concentration and the solvent conditions: it may either form the native state; it may associate with another chain to form a domain-swapped dimer; or several chains may aggregate or form a fibril



aggregation (see *Figure 10-7*). The exchange of domains between identical molecules is actually a special case of a range of phenomena termed "domain swapping" (Bennett et al., 1994). The term describes situations where two or more proteins exchange part of their structure (not necessarily whole domains) to form intertwined oligomers (Rousseau et al., 2003). Domain swapping has been implicated in amyloid fibril formation, although not all models of amyloid formation involve domain swapping (Nelson and Eisenberg, 2006).

9.2 Multi-Subunit Proteins

Almost all large proteins are formed by the association of subunits. There are several evolutionary advantages to forming multimers: regulation of catalytic activity through cooperativity between the subunits, generation of new functions and enzymatic activities, formation of large structures, increasing

protein stability, and facilitating the folding process. Subunit–subunit interfaces are very diverse regarding the nature and distribution of inter-subunit interactions. About one-third of the interfaces have a large and contiguous hydrophobic patch surrounded by a ring of inter-subunit polar interactions; the remaining two thirds show a mixture of small hydrophobic patches, polar interactions, and water molecules scattered over the interface area.

There are several experimental techniques to study the folding and association of multi-subunit proteins. Depending on the strength of the binding between the subunits, the experimentalist can attempt to produce folded but dissociated subunits under equilibrium conditions, using techniques such as dilution, cold dissociation, chemical modification, ligand-induced dissociation, mildly denaturing conditions, or elevated pressure (Jaenicke and Lilie, 2000). If unfolded monomers cannot be obtained, folding and association should be studied as coupled processes. Several methods can be used to monitor the association state of the protein after starting reassociation of the (folded or unfolded) subunits. By rapid chemical cross-linking, snapshots can be taken during the reassociation process and investigated by gel electrophoresis. Other methods include hybridization with isoenzymes or modified subunits and measurement of relative reactivation. There is also a wide range of biophysical methods to monitor and/or measure the thermodynamic parameters of protein–protein interactions (Lakey and Raggett, 1998), including surface plasmon resonance, isothermal titration calorimetry, fluorescence energy transfer, mass spectrometry, light scattering, and high-pressure liquid chromatography. After collecting the time-dependent data, a reaction equation can be fit and a reaction scheme established. The assembly of multi-subunit proteins can usually be described as a series of unimolecular (isomerization) and bimolecular (association) steps

$$4M_u \rightarrow 4M \rightarrow 2D' \rightarrow 2D \rightarrow T' \rightarrow T,$$

where M represents a monomeric, D a dimeric, and T a tetrameric state, and the subscript "u" represents the unfolded state.

Are folding and association separate events, or are they more or less coupled? The traditional view is that monomers assume a near-native conformation before binding to their partners. In recent years, however, several proteins have been found that are intrinsically unstructured as monomers and only fold upon binding to DNA, RNA, a membrane, or another protein (Dyson and Wright, 2002). Homodimers whose monomers only fully fold upon dimerization include troponin C site III (Monera et al., 1992), Arc repressor (Robinson and Sauer, 1996), FIS (factor for inversion stimulation) (Hobart et al., 2002), Trp repressor (Gloss et al., 2001), and the dimeric form of p53 (Mateu et al., 1999). The HIV gp41 protein is a homotrimer with intrinsically unfolded monomers (Marti et al., 2004). It has been shown that the binding mechanism (whether monomer folding is coupled to binding or there are folded monomeric intermediates) is determined by the native topology, especially the number of inter- and intramolecular contacts and the hydrophobicity of the interface (Levy et al., 2004). Flexibility of the chain seems to play a major role in the binding mechanism (Levy et al., 2005); one important manifestation of this is the fly-casting effect (Shoemaker et al., 2000): a relatively unstructured chain can have a greater "capture radius" to "catch" its binding partner before fully folding.

10 Protein Folding in the Cell

Under carefully chosen in vitro conditions, small single-domain proteins fold in a cooperative and reversible manner. Usually, the experiment is carried out in dilute solutions at low temperatures where the folding reaction is not complicated by off-pathway reactions such as aggregation. Contrarily, the interior of a cell is a highly crowded environment with an estimated protein concentration of 300 mg/ml (Zimmerman and Trach, 1991), and in extreme cases, such as in a thermophilic archaeon, the temperature can exceed 80°C. Another difference is that in vivo the folding process is not separated from the relatively slow synthesis of the polypeptide chain (Creighton, 1990; Jaenicke, 1991). Folding cannot complete before a folding unit is completely synthesized. Nascent chains emerging from the ribosome should avoid formation of misfolded intermediates and aggregation. To ensure efficient folding, cells have evolved a large and

diverse group of proteins that assist the formation of the native structure. These proteins form the complex machinery of "molecular chaperones," whose function is to guide other proteins to their proper folding and unfolding routes and help the assembly or disassembly of macromolecular structures, without becoming permanent components of these structures. Some chaperones are also stress or heat-shock proteins because the need for chaperone function increases under conditions of stress that cause proteins to unfold or misassemble (Ellis, 2005).

Although chaperone systems in eukaryotic cells are more complex than in prokaryotes, several homologous classes of chaperones have been identified in these different cell types. Two major classes of chaperones, the Hsp70s and the cylindrical chaperonin complexes, protect nonnative polypeptide chains from misfolding and aggregation in the cytosol of prokaryotes and eukaryotes, and inside the chloroplasts and mitochondria (Ostermann et al., 1989; Hartl, 1996), in an ATP-dependent manner. The two classes exhibit distinct structural and functional properties and show entirely different mechanisms of action. Other protein classes, such as the Hsp40s, nucleotide exchange factors, or ADP-destabilizing factors, act as cofactors to the chaperones. Here, we briefly summarize what is known about the major chaperone classes.

10.1 The Hsp70 Family

Hsp70s are distributed in all types of cells and in all cellular compartments. The Hsp70 family is a highly conserved protein family consisting of numerous homologs with distinct cellular functions (Frydman, 2001). These chaperones have a molecular mass of approximately 70 kDa. The molecules consist of two domains, the 44 kDa N-terminal domain, which mediates ATP binding (Flaherty et al., 1990) and the small C-terminal domain, which binds the protein substrate (Zhu et al., 1996). Substrate binding and release is modulated by ATP binding and hydrolysis. When ATP is bound, substrate binding and release occur rapidly, while with ADP bound, both substrate binding and release are slow. The function of Hsp70s is tuned by cofactors modulating substrate and nucleotide binding. Members of the Hsp70 family are DnaK in bacteria, Ssa and Ssb in yeast, and Hsc70 and Hsp70 in mammals.

Hsp70s have a substrate-binding cleft, which recognizes extended stretches of polypeptide chains rich in hydrophobic residues (Flynn et al., 1991; Rudiger et al., 1997), such as segments of partially unfolded proteins in nonnative conformations. Bound substrates are protected from aggregation; however, folding is obviously not possible in the bound state. Repeated binding and release might keep substrate proteins in extended, monomeric conformation giving them the chance to assume their native structure. However, many of the Hsp70-bound substrates are transferred to the real "folding machine," the chaperonin system.

Hsp40s are cofactors that stimulate ATP hydrolysis by Hsp70s. Hsp40s are also capable of binding substrates and can pass the bound substrates on to a Hsp70 molecule. The molecule usually consists of an N-terminal J-domain, which is responsible for Hsp70-binding and a C-terminal chaperone domain containing hydrophobic patches for substrate binding (Sha et al., 2000). Representatives of the family are DnaJ in *E. coli*, Ydj1 and Sis1 (ribosome associated) in yeast, and Hdj1–2 and Hsp40 in mammals. Several eukaryotic homologs contain only a J-domain, which may have a role in the localization of Hsp70s.

Nucleotide exchange factors promote the release of ADP from members of the Hsp70 family. This group includes GrpE, a 23-kDa protein in *E.coli*, and its homologs in mitochondria and chloroplasts, as well as several 60-kDa proteins including Sti1 in yeast and Hop in mammals. In mammals, the Bag1 protein (Hohfeld and Jentsch, 1997; Takayama et al., 1997) also promotes the release of ADP from Hsp70 and the subsequent substrate release. These are modular proteins with domains that might have a role in connecting different chaperone systems. Hop may link Hsp70 to the Hsp90 system (Gross and Hessefort, 1996), and Bag1 contains an ubiquitin homology domain, suggesting a possible direction of Hsp70-bound substrates to the S26 proteasome system (Luders et al., 2000; Terada and Mori, 2000).

ADP stabilizing factors, such as the 48-kDa Hip protein in mammals, bind Hsp70 and stabilize the ADP–Hsp70 complex.

Small chaperones exhibiting Hsp70-like activity have a function similar to that of Hsp70, but unlike Hsp70, their function is independent from nucleotide binding. The trigger factor (TF) in *E. coli* has an overlapping function with DnaK. Additionally, it binds to the ribosome and displays prolyl-isomerase activity (Hesterkamp et al., 1996).

Hsp70 can be functionally substituted by the ubiquitous prefoldin (GimC) chaperone. Prefoldin is a heterohexamer that can interact with nascent polypeptides in vitro. The crystal structure of prefoldin reveals a unique quaternary structure forming a novel class of chaperones (Siegert et al., 2000).

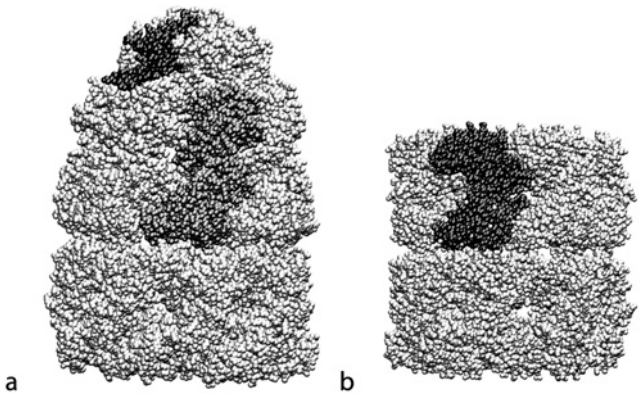
10.2 The Folding Cage of Chaperonins (Hsp60 Family)

Chaperonins are barrel-like multi-subunit complexes that primarily promote ATP-dependent protein folding (Farr et al., 2000; Brinker et al., 2001). The unique mechanism of chaperonins involves the capture and isolation of substrate polypeptide chains inside the chamber of the complex.

AU3

In the case of group I chaperonins such as GroEL in *E. coli*, the mitochondrial Hsp60, and the Rubisco-binding subunit (RBP) in plants, the system requires a Hsp10-type co-chaperonin that acts as a "cap" or "lid." The GroEL system with GroES as co-chaperonin is capable of correctly folding proteins of sizes up to 60–70 kDa and with multiple domains (Houry et al., 1999). GroEL is a homooligomer of 14 subunits (see *Figure 10-8*). These subunits consist of an apical, an intermediate, and an equatorial domain and are arranged in two stacked rings forming two chambers. The apical domains in the open conformation

Figure 10-8
(a) The structure of the GroEL complex with the GroES lid on (closed state). (b) The structure of the GroEL complex in the open state (no GroES lid present). One subunit of both GroEL and GroES is shown in a darker color



provide hydrophobic side chains that can interact nonspecifically with the exposed hydrophobic surface of the unfolded substrate chain. Subsequent binding of the GroES cap and seven ATP molecules to GroEL triggers a conformational change resulting in an increased volume of the central cavity, a separation of the hydrophobic residues, and an exposure of hydrophilic residues in the apical region (Shtilerman et al., 1999). This may induce partial unfolding of the substrate molecule and its release into the inside of the central cavity of the chaperonin. The isolated environment of the central cavity is ideal for the substrate molecule for folding up into its native structure. Upon completion of ATP hydrolysis, which may take approximately 10 s, binding of seven new ATP molecules to the *trans* ring triggers the dissociation of the GroES cap and the substrate molecule is released. In the case of incomplete folding, the substrate molecule can be recaptured and the "annealing" and relaxation cycle can be repeated until the molecule correctly folds.

Group II chaperonins are found in eukaryotes and in archaea and are homologous to group I chaperonins with a sequence identity up to 40%, showing similar double-ring architecture. The eukaryotic group II chaperonin named TCP-1 (for tailless complex polypeptide-1) or CCP (for chaperonin-containing TCP-1) is a heterooctamer consisting different subunits of 55–60 kDa (Kubota et al., 1995). Group II chaperonins have a helical protrusion on the apical domain that takes the place of the co-chaperonin GroES. The crystal structure of the thermosome from *Thermoplasma acidophilum* reveals that this protrusion can assume different conformations in the open state including α-helix and β-sheet, which can increase the plasticity for binding different types of substrates.

10.3 The Hsp90 Chaperone System

Hsp90 plays a central role in eukaryotes in the regulation of the components of signal transduction systems such as tyrosine kinases and steroid hormone receptors. Hsp90 is ATP dependent and requires interaction with several cofactors, some of them having chaperone activity. An example is the substrate transfer from Hsp70 to Hsp90, which is facilitated by Hop (Hsp70–Hsp90-organizing protein). Hop contains binding sites for both chaperones and contributes to the rapid transfer of receptor molecules from Hsp70 to Hsp90. A detailed review on Hsp90 function has been published recently (Pearl and Prodromou, 2006).

10.4 Chaperone-Assisted Assembly of Cellular Complexes

Recent work has shown the role of nuclear chaperones in the assembly of nucleosomes and has led to the discovery of a cytosolic chaperone required for mammalian proteasome assembly, suggesting that besides the folding of individual proteins, the formation of oligomeric complexes may also be assisted by chaperones (Ellis, 2006).

10.5 Cooperation Between the Different Chaperone Systems

Newly synthesized polypeptide chains, proteins losing their native state upon stress conditions, and other proteins requiring translocation in the cell may interact with several different chaperone systems. Chaperones having overlapping functions may compete for their substrate molecules as well. Hsp70 binds polypeptides and prevents aggregation. This may be sufficient for the correct folding of some proteins. Others may require further transfer from Hsp70 to the more efficient folding machine: the Hsp60 system. In *E. coli*, the cooperation between DnaK and GroEL has been proven in vivo (Teter et al., 1999). TF, which may substitute for the function of DnaK, also appears to cooperate with GroEL in substrate binding (Kandror et al., 1997). In eukaryotic cells, Hsp70 and TCP-1 associate, indicating a close functional relationship (Lewis et al., 1992). As mentioned above, Hsp70 and Hsp90 may also interact in receptor transfer.

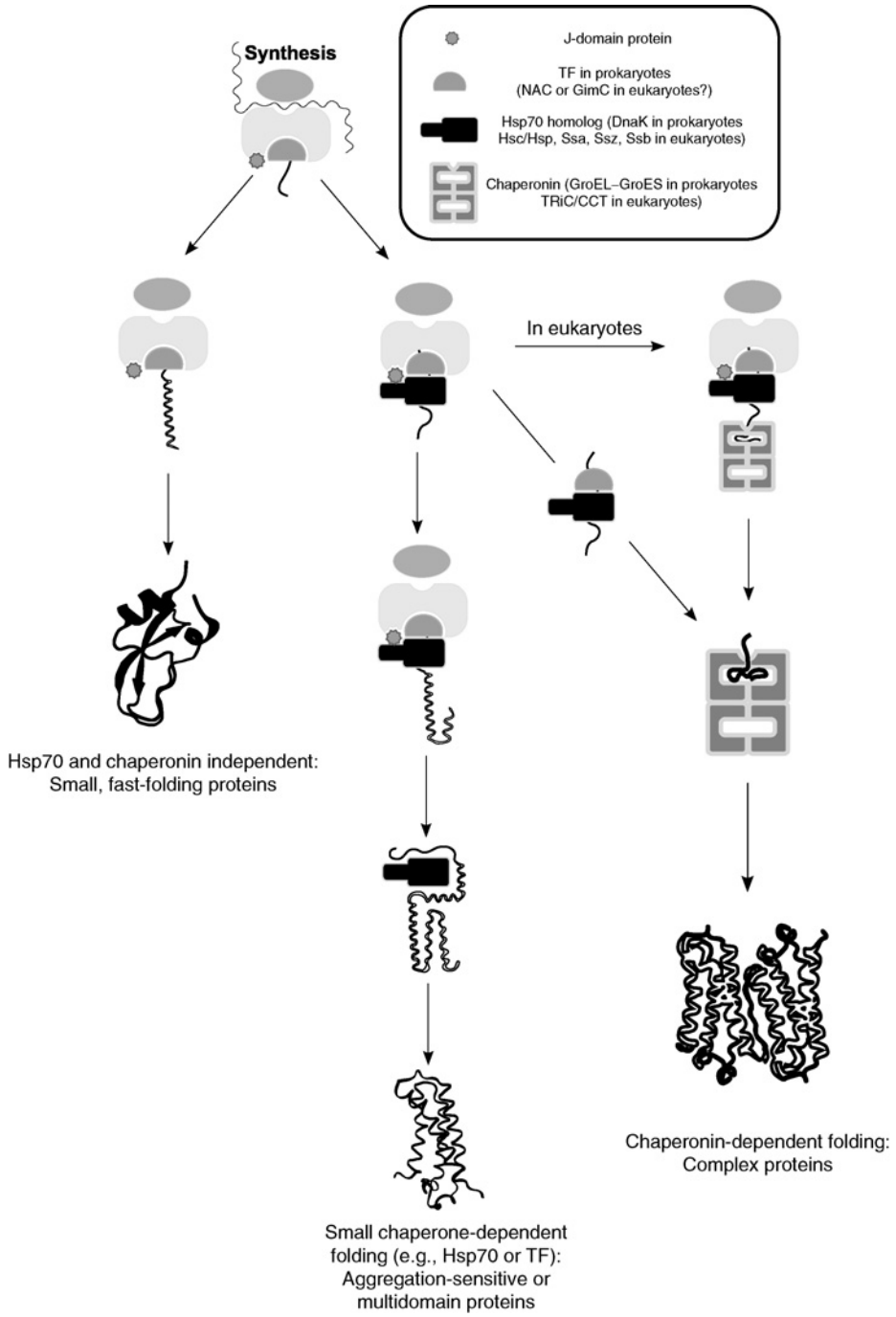
Figure 10-9 shows a schematic representation of the general view of de novo protein folding in the cytosol.

10.6 Chaperones and Neurodegenerative Diseases

The pathogenesis of several neurodegenerative diseases is associated with protein misfolding, aggregation, and deposition of the protein, which may be manifested in cell degeneration and loss of function of the affected cells or organs. These degenerative disorders include polyglutamine (polyQ) tract diseases, Alzheimer's and Parkinson's disease, amyotrophic lateral sclerosis, and Creutzfeldt–Jakob syndrome (Chiti and Dobson, 2006). Chaperones whose function is to prevent protein aggregation in the cell may play a crucial role in the onset or progress of neurodegenerative diseases.

Figure 10-9

A schematic representation of the general view of de novo protein folding in the cytosol



11 Protein Misfolding and Aggregation

Native states of proteins almost always represent the thermodynamically most stable conformation under physiological conditions (Vendruscolo et al., 2003). All the information regarding the native structure is hidden in the amino acid sequence. However, as we have seen, correct folding is a challenge for proteins in a living cell and only a part of the proteins can assume their native structure spontaneously. In the crowded milieu of the cell, efficient protein folding and transport depend on the presence of a complex machinery of chaperones, chaperonins, and cofactors. The primary mission of this machinery is to prevent the aggregation of nascent polypeptide chains and proteins that unfold upon environmental stress (Frydman, 2001).

The failure of a specific protein to adopt or maintain its native functional conformation may result in pathological conditions referred to as "protein misfolding diseases." Misfolding diseases include a wide range of diseases with different pathological mechanisms. The loss of the normal cellular function because of a reduction in the number of functional protein molecules is responsible for diseases such as cystic fibrosis (Amaral, 2004) and early-onset emphysema (Lomas and Carrell, 2002). The major group of misfolding diseases, however, is associated with aggregation and deposition of proteins in the human body in the form of organized, fibrillar aggregates, generally termed amyloid fibrils. On one hand, unwanted protein aggregation may be caused by malfunctioning of the cellular protein quality-control machinery. It may occur when the ubiquitin-proteasome protein degradation system cannot eliminate misfolded, aggregation prone molecules (Ross and Pickart, 2004; Mandel et al., 2005); when the chaperone machinery performs insufficiently (Lee and Tsai, 2005); if the normal cellular transport route of a protein is damaged; or when inappropriate protease activity produces amyloidogenic protein fragments. On the other hand, aggregation and amyloid formation of a protein may be promoted by an increased expression level or by pathological mutations destabilizing the structure and inducing intermediate amyloidogenic conformations (Chiti and Dobson, 2006).

11.1 Degenerative Diseases Associated with Amyloid Deposition

Amyloid deposition is associated with more than 20 human degenerative diseases. We distinguish different groups of diseases by the location of amyloid deposits in the body. Neurodegenerative conditions such as Alzheimer's, Parkinson's, and Huntington's disease as well as spongiform encephalopathies affect the central nervous system. In nonneuropathic localized amyloidoses, protein deposition occurs in a certain type of tissue such as Langerhans' islands in type II diabetes. In systemic amyloidoses such as AL amyloidosis, which involves the deposition of immunoglobulin light-chain fragments, protein deposition is not limited to a single tissue. Alzheimer's and Parkinson's disease are sporadic and usually develop with aging, which suggests the role of the protein quality-control machinery in the disease, although hereditary forms are also documented. Other conditions are hereditary, arising from specific mutations, such as lysozyme and fibrinogen amyloidosis. The special property of spongiform encephalopathy is that it can be transmissible in humans and mammals (Chiti and Dobson, 2006). Hemodialysis-related amyloidosis represents the first amyloid disease that is a complication of a medical therapy (Gejyo et al., 1985).

Protein aggregates can accumulate both extracellularly and intracellularly. The term "intracellular inclusion" has been suggested as more appropriate for amyloid-like aggregates depositing inside the cell (Westermark et al., 2005). However, herein we use the term "amyloid fibril."

11.2 The Structure and Morphology of Amyloid Fibrils

A common characteristic feature of amyloid fibrils formed from different proteins is the well-ordered structure with high β-structure content. X-ray fiber diffraction studies have shown that the β-strands are oriented perpendicularly to the fibril axis (Sunde and Blake, 1997). Electron microscopy (EM) and atomic

force microscopy (AFM) revealed that amyloid fibrils possess diverse morphologies at the fibrillar level. Single protofilaments can be straight or curved, with a diameter of 2–5 nm, showing no helical twist. Fibrils usually consist of 2–6 protofilaments, twisting together in a rope-like or ribbon form with a diameter of 7–15 nm (Serpell et al., 2000).

High-resolution structure determination of amyloid fibrils is a grand challenge. Traditional spectroscopic methods fail because of the insoluble nature or the large size of the fibrils. Crystallization for X-ray is complicated because fibrils favor growth in one direction. The crystal structure of the amyloid form of the GNNQQNY peptide has recently been solved (Nelson and Eisenberg, 2006). An extended β-sheet is formed in the crystal, with each of the β-strands consisting of a single peptide. The structure revealed a tight packing of the side chains between two β-sheets, excluding water molecules.

Solid-state NMR spectroscopy may provide high-resolution information on amyloid structure. Tycko (2006) and coworkers built a model structure of the Aβ(1–40) amyloid β peptide, associated with Alzheimer's disease, based on solid-state NMR constraints. In the model, the different Aβ molecules are stacked on each other in a parallel arrangement and in register, forming two β-sheets. Every single Aβ molecule contributes to two β-strands, one in each β-sheet.

Without providing detailed structural information, hydrogen–deuterium exchange methods combined with NMR spectroscopy, as well as limited proteolysis with mass spectrometry, are capable of collecting site-specific information on the extent of the rigid amyloid core (Hoshino et al., 2002; Myers et al., 2006a).

The morphology of amyloid fibrils grown in vitro highly depends on solution conditions such as buffer composition, pH, temperature, and protein concentration. Even under the same conditions, fibrils with different morphologies can be formed from the same polypeptide. Structural studies revealed that this polymorphism of amyloid fibrils is a reflection of different underlying structures at the molecular level (Petkova et al., 2005).

AFM and EM have shown a wide variety of protein aggregates depending on the conditions. Proteins may form disordered aggregates, oligomers, spherical aggregates, prefibrillar aggregates, and fibrils with various morphologies (Kad et al., 2003; Stine et al., 2003). ❯ *Figure 10-10* shows a unified view of the various types of structures that can be formed by polypeptide chains in vivo or in vitro.

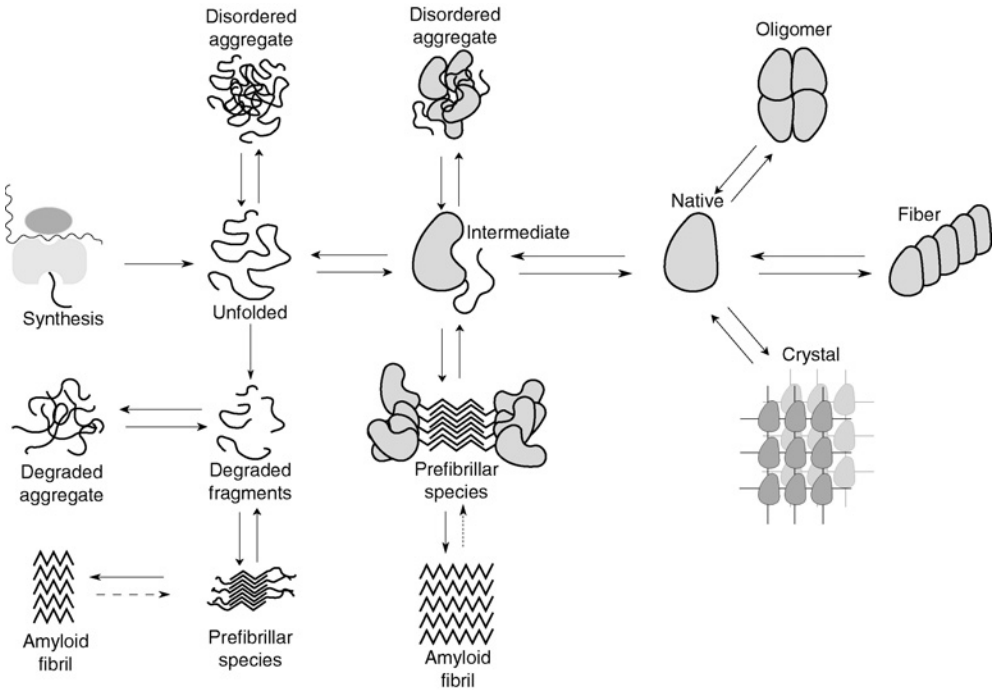
11.3 Mechanism of Amyloid Fibril Formation

The kinetics of amyloid formation and the amyloid content of a protein solution can be studied by light scattering, thioflavin T fluorescence, or other spectroscopic methods. Amyloid formation is usually a nucleation-dependent reaction showing an initial lag phase, followed by rapid growth. Clearing the solution of any aggregated material by ultracentrifugation prior to the reaction may significantly increase the lag phase. Oppositely, addition of a small amount of preformed aggregated material can reduce or eliminate the lag phase. These observations suggest that the lag phase is the time required for proper nucleation. Under some conditions, no lag time is observable, suggesting that nucleation is not always the rate-limiting step (Uversky et al., 2002; Pedersen et al., 2004). To understand the in vivo mechanism of amyloidoses, it is important to understand the nucleation process and carefully examine the reaction during the lag phase. Oligomers, spherical or chain-like aggregates, sometimes termed "protofibrils," forming prior to fibril formation have been observed in many systems (Chiti and Dobson, 2006). It is extremely important to understand the mechanism of the formation of these prefibrillar species. Cell culture and in vivo studies have revealed their toxicity for living cells, and they may be involved in the pathogenesis of diseases such as Alzheimer's (Lue et al., 1999; McLean et al., 1999).

11.4 Amyloid Formation of Globular Proteins

One group of amyloid-forming systems consists of peptides or protein fragments that are natively unfolded in the monomeric form (Alzheimer's Aβ, Parkinson's α-synuclein). It is generally accepted that stable globular proteins need to partially unfold to an amyloidogenic intermediate for fibril formation. This is

Figure 10-10
A unified view of the major types of structure that can be formed by polypeptide chains



supported by experimental data showing increased amyloidogenicity under conditions that destabilize the native state. Destabilizing mutations may also promote amyloid formation, which explains the mechanism of some hereditary diseases (Raffen et al., 1999; Canet et al., 2002).

Moreover, under carefully chosen denaturing conditions, most proteins are capable of aggregation and amyloid formation in vitro, suggesting that the amyloid state is a general property of polypeptide chains (Stefani and Dobson, 2003). However, out of thousands of different proteins in the human body, only some two dozen are responsible for diseases, indicating that protein sequences and the cellular machinery have evolved to avoid unwanted protein aggregation.

11.5 Physicochemical and Sequence Determinants of Amyloid Formation

Hydrophobicity of the peptide chain has been shown to influence its aggregation propensity (Otzen et al., 2000). Clusters of consecutive hydrophobic residues are avoided by evolution (Schwartz et al., 2001). Another crucial factor is the charge of the polypeptide chain. A high net charge may prevent aggregation of the polypeptide (Chiti et al., 2002). Secondary structure propensity may also affect amyloid formation: a high propensity to form β-sheet structure and low α-helix propensity are likely to increase the probability of aggregation (Chiti and Dobson, 2006). Alternating patterns of polar and nonpolar residues, which promote β-sheet formation, are less frequent in natural protein sequences than expected on a random basis (Broome and Hecht, 2000). Proline residues break β-sheet structures, and hence inhibit aggregation (Steward et al., 2002; Parrini et al., 2005). The presence of β-bulges on β-strands exposed on the protein surface has been found to protect against the formation of intermolecular β-sheets (Richardson and Richardson, 2002).

11.6 Factors Inducing or Inhibiting Amyloid Formation Under Physiological Conditions

Amyloid formation may be facilitated under proper in vitro conditions, but in vivo, it always occurs under physiological conditions. Some disease-related proteins such as β_2-microglobulin cannot form amyloid fibrils at physiological pH in vitro, suggesting the presence of unknown factors contributing to the aggregation process in vivo. Collagen, apolipoprotein-E, heparin, serum amyloid P component, and low concentrations of sodium dodecyl sulfate have been reported to promote the aggregation of β_2-microglobulin under physiological conditions (Yamamoto et al., 2004; Myers et al., 2006b). Other molecules binding to and stabilizing the native state may prevent oligomerization or amyloid formation of disease-related proteins in the human body, and might serve as effective therapeutic agents in the future.

12 New Biophysical Techniques for the Study of Protein Folding

In the last decade a wide range of physical and chemical methods complemented the fundamental techniques to study protein folding. We give a brief summary of the advances of these methods.

12.1 Rapid Mixing Methods

The mechanism of protein folding can be studied by two different groups of approaches. Equilibrium methods provide information about possible folding intermediate states or deduce rate constants from the molecular fluctuations or dynamic properties of the system. Relaxation methods follow the change of the system evolving toward a new equilibrium after a rapid perturbation of its extrinsic variables, such as temperature, pH, pressure, or solvent composition (Roder et al., 2004). The time required for folding varies greatly among proteins, ranging from microseconds to minutes (Roder and Shastry, 1999; Roder et al., 2006). The smallest protein molecules with no folding intermediates fold on the microsecond timescale, which, on one hand, might make them suitable for in silico-folding simulation studies. On the other hand, such fast reactions make the experimental detection of events during the folding process difficult. Basic techniques in the study of folding kinetics are stopped-flow fluorescence spectroscopy and stopped-flow circular dichroism. These techniques are capable of monitoring the formation of secondary and tertiary structures during the folding reaction with millisecond time resolution. In such experiments, rapid processes occurring within the dead time of the measurement were observed for many proteins. The challenge to resolve this initial burst phase and to reveal the structural changes taking place during the first millisecond stimulated the development of new, rapid kinetic techniques capable of triggering and monitoring the folding process on the submillisecond timescale. While the conventional stopped-flow apparatus, in which a small volume of a freshly made mixture containing the reacting components is injected into the measurement cell, is quite economical and offers a wide range of applications (Gibson and Milnes, 1964), its dead time is usually about 1 millisecond or longer. Continuous-flow methods extend the time resolution to the microsecond time range (Chan et al., 1997; Takahashi et al., 1997; Shastry et al., 1998; Akiyama et al., 2002). In the continuous-flow cell, solutions are mixed under highly turbulent conditions to achieve complete mixing. The kinetics of the reaction is monitored under steady-state flow conditions as a function of the distance downstream from the mixer by using relatively simple and inexpensive detection methods. Using this technique, it has become possible to study the initial collapse and formation of intermediates in the early stage of the folding reaction, during the burst phase (Uzawa et al., 2004; Welker et al., 2004; Kimura et al., 2005).

12.2 Real-Time NMR Spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy has greatly contributed to our understanding of the protein-folding problem. NMR provides high spatial resolution and a broad timescale ranging from

picoseconds to days. Hydrogen–deuterium exchange experiments, revealing dynamical events at an atomic level, have illuminated the process of unfolding from the native state and the structure of folding intermediates (Wagner and Wuthrich, 1982; Wand et al., 1986; Bai et al., 1995). The quenched-flow pulse-labeling technique has enabled researchers to study the early stages of protein folding using a conventional NMR instrument (Roder et al., 1988; Udgaonkar and Baldwin, 1988; Radford et al., 1992). NMR studies have characterized the properties of denaturant-induced equilibrium folding intermediate states such as the molten globule (Arai and Kuwajima, 2000). In equilibrium systems, the rates of conversion between distinct conformational states can be calculated from a line shape analysis of the NMR resonances (Huang and Oas, 1995), and therefore can provide kinetic data on folding.

Slow folding reactions such as *cis–trans* prolyl isomerization can be directly followed by sequential recording of one-dimensional (1D) NMR spectra (Balbach et al., 1999). This method is particularly useful for discovering intermediates formed at the late stages of the folding process. Using a stopped-flow device for injection of the protein solution into the NMR tube that already contains the denaturant or the refolding buffer pushes the dead time of mixing below 1 s (Zeeb and Balbach, 2004). One of the first proteins studied by real-time NMR was α-lactalbumin (Balbach et al., 1995). 1D-NOE (nuclear Overhauser effect) experiments revealed the native-like compactness of the transient molten globule state of α-lactalbumin (Forge et al., 1999). These experiments also demonstrated that the transient intermediate closely resembles the well-characterized stable molten globule state formed at low pH. While 1D-NMR spectra have limited resolution, multidimensional NMR can provide high spatial resolution information on the folding process. Because recording multidimensional spectra is time consuming, only slow processes could be followed directly by sequential recording (Liu et al., 1996). Balbach and coworkers (1996, 1999) developed new methods to reconstruct the kinetic history of folding reactions from a single two-dimensional NMR spectrum recorded during the entire time course of the reaction. The basis of these methods is that the line widths and intensities reflect the history of the folding events occurring during spectral accumulation. When applied to α-lactalbumin, the technique demonstrated the cooperative nature of the folding of the main chain.

12.3 Chemically Induced Nuclear Polarization

Chemically induced nuclear polarization (CIDNP) can be used to probe the solvent accessibility of certain aromatic residues in proteins (Mok et al., 2003; Mok and Hore, 2004). The reactive collision of polarizable amino acids such as tryptophan, tyrosine, and histidine with a photoexcited dye such as flavin mononucleotide (FMN) results in an electron transfer (in the case of Trp and Tyr) or proton transfer (His) reaction forming a pair of radicals. Electron-nuclear hyperfine interactions between the two radicals result in a significant enhancement of NMR signals. The "photosensitizer" flavin molecule can be excited by laser as light source. For the photoreaction to take place, the aromatic side chains must be accessible to the photosensitizer, e.g., located on the surface of the protein molecule. The CIDNP spectrum is recorded immediately after the laser flash and corrected by a "dark" spectrum recorded without irradiation. Besides the equilibrium studies of protein surfaces, the technique can be combined with a stopped-flow apparatus and in this way it can be used to study folding intermediates. Using CIDNP pulse-labeling technique, the exposed tryptophan and tyrosine residues in a molten globule state can be identified (Lyon et al., 2002; Mok et al., 2003).

12.4 High-Pressure NMR Spectroscopy

When high pressure is applied to a protein solution, it shifts the conformational equilibrium of the protein molecules toward lower volume conformers, thereby decreasing the partial molar volume of the protein. The combination of high pressure with heteronuclear two-dimensional NMR spectroscopy provides atomic resolution information on the structure of the protein molecule at different stages of the folding process (Kamatari et al., 2004). By varying the pressure, one can explore the conformational space from the folded to the unfolded conformer. In recent years, numerous studies using high-pressure NMR spectroscopy have

been carried out (Akasaka and Yamada, 2001) on locally disordered (Kuwata et al., 2001; Kuwata et al., 2002; Kitahara et al., 2005), molten globule (Kitahara et al., 2002; Lassalle et al., 2003), unfolded (Kamatari et al., 2001; Arnold et al., 2002; Refaee et al., 2003) as well as oligomeric or aggregated states of proteins (Niraula et al., 2004; Silva et al., 2006).

12.5 Protein Folding and Dynamics Studied by Mass Spectrometry

Mass spectrometry of protein molecules has become a rapidly developing field in the last decade (Konermann and Simmons, 2003; Eyles and Kaltashov, 2004). In comparison with NMR spectroscopy, which provides site-specific information averaged in time, mass spectrometry is capable of detecting different conformers coexisting in the protein solution. This method is especially useful for the study of low-populated intermediate states and is free of the molecular size limitation of NMR spectroscopy. Because of its high sensitivity, a protein concentration in the femtomolar range is sufficient for analysis. Structural and dynamic properties of various conformational states can be studied by hydrogen/deuterium exchange (HDX) combined with mass spectrometry. Recently, Kaltashov and coworkers investigated the conformational ensemble of the molten globule state of ubiquitin (Hoerner et al., 2005). Using protein ion fragmentation in the gas phase, they evaluated the stability of various segments of the protein in the molten globular state. By the method of pulse-labeling HDX-MS, it is possible to study the kinetics of folding and to explore complex folding scenarios with parallel pathways (Konermann and Simmons, 2003). Co-populated protein conformers can be detected and characterized directly by electrospray ionization mass spectrometry (ESIMS) (Mohimen et al., 2003; Borysik et al., 2004). Protein surface areas in solution may be determined by ESIMS (Kaltashov and Mohimen, 2005).

Limited proteolysis with ESIMS provides site-specific structural information on different conformational states of the protein molecules including protein aggregates and the amyloid state (Myers et al., 2006a).

12.6 Mechanical Unfolding of Proteins

In the first studies of the mechanical unfolding of single protein molecules using AFM, the giant sarcomeric protein titin, consisting of a large number of immunoglobulin segments, was used (Erickson, 1997; Rief et al., 1997; Rounsevell et al., 2004). Because of the heterogeneity of titin domains, it was not possible to assign the individual force peaks to specific domains. Using tandem repeats of a single domain, constructed by protein engineering techniques, it was possible to explain the mechanical characteristics of single domains in terms of their specific structures (Carrion-Vazquez et al., 1999; Carrion-Vazquez et al., 2000).

Using force-measuring optical tweezers, it is possible to induce mechanical unfolding and refolding of individual molecules (Kellermayer et al., 1997). In a recent work, Cecconi and coworkers (2005) showed that *E. coli* ribonuclease H molecule unfolds in a two-state manner and refolds through a transient molten globule-like intermediate.

We may expect significant progress in the application of other new techniques such as the study of single-molecule folding kinetics by optical techniques in the near future (Lipman et al., 2003).

13 Conclusions

In spite of the great advances in experimental technique and the tremendous boost in computational power we have witnessed in the past decades, the protein-folding problem is still far from solved. We already have a good understanding of some of the simpler protein-folding mechanisms, and we can simulate the folding of a few, very small proteins. However, we are still at a loss when the folding behavior of more complex, multidomain proteins is to be explained, especially when protein–protein interactions, misfolding, aggregation, and amyloid formation complicate the situation—such as in a living cell. Although we have known

since Anfinsen that sequence determines structure (in a given environment), and the field of protein structure prediction has made great progress, we are still not, in general, capable of predicting the structure from a sequence. The research of protein folding, however, is no longer an exotic field with only theoretical significance. Diseases such as Alzheimer's or Parkinson's have taught us that protein folding can go awry, and misfolded or aggregated proteins can cause a lot of trouble. In order to successfully defend us against these and other "folding diseases," we should reach a level of understanding of folding phenomena that is not just descriptive but allows us to influence the folding behavior of proteins in the way we want. Therefore, the research of folding continues, and in all likelihood will bring about great advances in the development of new experimental techniques and new theoretical approaches.

Acknowledgments

A.S. was supported by a Bolyai János fellowship. A.S. and P.Z. were supported by grants from the Hungarian Scientific Research Fund (OTKA T046423 and NI-61915). S.O. was supported by a grant from the Hungarian Scientific Research Fund (OTKA D-38480).

References

Akasaka K, Yamada H. 2001. On-line cell high-pressure nuclear magnetic resonance technique: Application to protein studies. Methods Enzymol 338: 134-158.

Akiyama S, Takahashi S, Kimura T, Ishimori K, Morishima I, et al. 2002. Conformational landscape of cytochrome *c* folding studied by microsecond-resolved small-angle x-ray scattering. Proc Natl Acad Sci USA 99: 1329-1334.

Alm E, Baker D. 1999. Prediction of protein-folding mechanisms from free-energy landscapes derived from native structures. Proc Natl Acad Sci USA 96: 11305-11310.

Alonso DO, Daggett V. 1998. Molecular dynamics simulations of hydrophobic collapse of ubiquitin. Protein Sci 7: 860-874.

Amaral MD. 2004. Cftr and chaperones: Processing and degradation. J Mol Neurosci 23: 41-48.

Anfinsen CB. 1973. Principles that govern the folding of protein chains. Science 181: 223-230.

Arai M, Kuwajima K. 2000. Role of the molten globule state in protein folding. Adv Protein Chem 53: 209-282.

Arnold MR, Kremer W, Ludemann HD, Kalbitzer HR. 2002. ^{1}H-NMR parameters of common amino acid residues measured in aqueous solutions of the linear tetrapeptides gly-gly-x-ala at pressures between 0.1 and 200 MPa. Biophys Chem 96: 129-140.

Bai Y, Sosnick TR, Mayne L, Englander SW. 1995. Protein folding intermediates: Native-state hydrogen exchange. Science 269: 192-197.

Baker D. 2000. A surprising simplicity to protein folding. Nature 405: 39-42.

Balbach J, Schmid FX. 2000. Proline isomerization and its catalysis in protein folding. Mechanisms of Protein Folding. Pain RH, editor. Oxford: Oxford University Press; pp. 212-249.

Balbach J, Forge V, Lau WS, van Nuland NA, Brew K, et al. 1996. Protein folding monitored at individual residues during a two-dimensional NMR experiment. Science 274: 1161-1163.

Balbach J, Forge V, van Nuland NA, Winder SL, Hore PJ, et al. 1995. Following protein folding in real time using NMR spectroscopy. Nat Struct Biol 2: 865-870.

Balbach J, Steegborn C, Schindler T, Schmid FX. 1999. A protein folding intermediate of ribonuclease T1 characterized at high resolution by 1D and 2D real-time NMR spectroscopy. J Mol Biol 285: 829-842.

Baldwin RL. 1989. How does protein folding get started? Trends Biochem Sci 14: 291-294.

Baldwin RL. 1996. On-pathway versus off-pathway folding intermediates. Fold Des 1: R1-R8.

Becker OM, Karplus M. 1997. The topology of multidimensional potential energy surfaces: Theory and application to peptide structure and kinetics. J Chem Phys 106: 1495-1517.

Bennett MJ, Choe S, Eisenberg D. 1994. Domain swapping: Entangling alliances between proteins. Proc Natl Acad Sci USA 91: 3127-3131.

Bergman LW, Kuehl WM. 1979. Formation of an intrachain disulfide bond on nascent immunoglobulin light chains. J Biol Chem 254: 8869-8876.

Blond S, Goldberg ME. 1986. Kinetic characterization of early intermediates in the folding of *E. coli* tryptophan-synthase β2 subunit. Proteins 1: 247-255.

Borysik AJH, Radford SE, Ashcroft AE. 2004. Co-populated conformational ensembles of β_2-microglobulin uncovered

quantitatively by electrospray ionization mass spectrometry. J Biol Chem 279: 27069-27077.

Braakman I, Helenius J, Helenius A. 1992. Manipulating disulfide bond formation and protein folding in the endoplasmic reticulum. EMBO J 11: 1717-1722.

Braakman I, Hoover-Litty H, Wagner KR, Helenius A. 1991. Folding of influenza hemagglutinin in the endoplasmic reticulum. J Cell Biol 114: 401-411.

Bradley P, Misura KMS, Baker D. 2005. Toward high-resolution de novo structure prediction for small proteins. Science 309: 1868-1871.

Brandts JF, Halvorson HR, Brennan M. 1975. Consideration of the possibility that the slow step in protein denaturation reactions is due to *cis-trans* isomerism of proline residues. Biochemistry 14: 4953-4963.

Brinker A, Pfeifer G, Kerner MJ, Naylor DJ, Hartl FU, et al. 2001. Dual function of protein confinement in chaperonin-assisted protein folding. Cell 107: 223-233.

Broome BM, Hecht MH. 2000. Nature disfavors sequences of alternating polar and non-polar amino acids: Implications for amyloidogenesis. J Mol Biol 296: 961-968.

Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG. 1995. Funnels, pathways, and the energy landscape of protein folding: A synthesis. Proteins 21: 167-195.

Bryngelson JD, Wolynes PG. 1987. Spin glasses and the statistical mechanics of protein folding. Proc Natl Acad Sci USA 84: 7524-7528.

Bulaj G. 2005. Formation of disulfide bonds in proteins and peptides. Biotechnol Adv 23: 87-92.

Bulaj G, Goldenberg DP. 2001. Φ-values for bpti folding intermediates and implications for transition state analysis. Nat Struct Biol 8: 326-330.

Bulleid NJ, Freedman RB. 1988. Defective co-translational formation of disulphide bonds in protein disulphide-isomerase-deficient microsomes. Nature 335: 649-651.

Canet D, Last AM, Tito P, Sunde M, Spencer A, et al. 2002. Local cooperativity in the unfolding of an amyloidogenic variant of human lysozyme. Nat Struct Biol 9: 308-315.

Capaldi AP, Kleanthous C, Radford SE. 2002. Im7 folding mechanism: Misfolding on a path to the native state. Nat Struct Biol 9: 209-216.

Carlsson U, Jonsson BH. 1995. Folding of β-sheet proteins. Curr Opin Struct Biol 5: 482-487.

Carra JH, Privalov PL. 1996. Thermodynamics of denaturation of staphylococcal nuclease mutants: An intermediate state in protein folding. FASEB J 10: 67-74.

Carrion-Vazquez M, Oberhauser AF, Fisher TE, Marszalek PE, Li H, et al. 2000. Mechanical design of proteins studied by single-molecule force spectroscopy and protein engineering. Prog Biophys Mol Biol 74: 63-91.

Carrion-Vazquez M, Oberhauser AF, Fowler SB, Marszalek PE, Broedel SE, et al. 1999. Mechanical and chemical unfolding of a single protein: A comparison. Proc Natl Acad Sci USA 96: 3694-3699.

Cecconi C, Shank EA, Bustamante C, Marqusee S. 2005. Direct observation of the three-state folding of a single protein molecule. Science 309: 2057-2060.

Chakraborty S, Ittah V, Bai P, Luo L, Haas E, et al. 2001. Structure and dynamics of the α-lactalbumin molten globule: Fluorescence studies using proteins containing a single tryptophan residue. Biochemistry 40: 7228-7238.

Chan CK, Hu Y, Takahashi S, Rousseau DL, Eaton WA, et al. 1997. Submillisecond protein folding kinetics studied by ultrarapid mixing. Proc Natl Acad Sci USA 94: 1779-1784.

Chan HS, Dill KA. 1996. A simple model of chaperonin-mediated protein folding. Proteins 24: 345-351.

Chan HS, Dill KA. 1998. Protein folding in the landscape perspective: Chevron plots and non-Arrhenius kinetics. Proteins 30: 2-33.

Chatrenet B, Chang JY. 1992. The folding of hirudin adopts a mechanism of trial and error. J Biol Chem 267: 3038-3043.

Cheng HN, Bovey FA. 1977. *Cis-trans* equilibrium and kinetic studies of acetyl-L-proline and glycyl-L-proline. Biopolymers 16: 1465-1472.

Chiti F, Dobson CM. 2006. Protein misfolding, functional amyloid, and human disease. Annu Rev Biochem 75: 333-366.

Chiti F, Calamai M, Taddei N, Stefani M, Ramponi G, et al. 2002. Studies of the aggregation of mutant proteins in vitro provide insights into the genetics of amyloid diseases. Proc Natl Acad Sci USA 99 (Suppl. 4): 16419-16426.

Chrunyk BA, Matthews CR. 1990. Role of diffusion in the folding of the α subunit of tryptophan synthase from *Escherichia coli*. Biochemistry 29: 2149-2154. AU5

Clementi C, Nymeyer H, Onuchic JN. 2000. Topological and energetic factors: What determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. J Mol Biol 298: 937-953.

Crane JC, Koepf EK, Kelly JW, Gruebele M. 2000. Mapping the transition state of the ww domain β-sheet. J Mol Biol 298: 283-292.

Creighton TE. 1979. Intermediates in the refolding of reduced ribonuclease A. J Mol Biol 129: 411-431.

Creighton TE. 1990. Protein folding. Biochem J 270: 1-16.

Creighton TE. 1997. Protein folding coupled to disulphide bond formation. Biol Chem 378: 731-744.

Creighton TE. 2000. Protein folding coupled to disulphide-bond formation. Mechanisms of protein folding. Pain RH, editor. Oxford: Oxford University Press; pp. 250-278.

Creighton TE, Goldenberg DP. 1984. Kinetic role of a metastable native-like two-disulphide species in the folding transition of bovine pancreatic trypsin inhibitor. J Mol Biol 179: 497-526.

Dabora JM, Pelton JG, Marqusee S. 1996. Structure of the acid state of *Escherichia coli* ribonuclease HI. Biochemistry 35: 11951-11958.

Daggett V, Fersht AR. 2000. Transition states in protein folding. Mechanisms of protein folding. Pain RH, editor. Oxford: Oxford University Press; pp. 175-211.

Daggett V, Fersht AR. 2003. Is there a unifying mechanism for protein folding? Trends Biochem Sci 28: 18-25.

Dautry-Varsat A, Garel JR. 1981. Independent folding regions in aspartokinase-homoserine dehydrogenase. Biochemistry 20: 1396-1401.

Day R, Daggett V. 2003. All-atom simulations of protein folding and unfolding. Adv Protein Chem 66: 373-403.

Demarest SJ, Horng JC, Raleigh DP. 2001. A protein dissection study demonstrates that two specific hydrophobic clusters play a key role in stabilizing the core structure of the molten globule state of human α-lactalbumin. Proteins 42: 237-242.

Dill KA. 1985. Theory for the folding and stability of globular proteins. Biochemistry 24: 1501-1509.

Dill KA. 1990. Dominant forces in protein folding. Biochemistry 29: 7133-7155.

Dill KA. 1999. Polymer principles and protein folding. Protein Sci 8: 1166-1180.

Dill KA, Chan HS. 1997. From levinthal to pathways to funnels. Nat Struct Biol 4: 10-19.

Dill KA, Bromberg S, Yue K, Fiebig KM, Yee DP, et al. 1995. Principles of protein folding—a perspective from simple exact models. Protein Sci 4: 561-602.

Ding F, Dokholyan NV. 2005. Simple but predictive protein models. Trends Biotechnol 23: 450-455.

Ding F, Buldyrev SV, Dokholyan NV. 2005a. Folding trp-cage to NMR resolution native structure using a coarse-grained protein model. Biophys J 88: 147-155.

Ding F, Guo W, Dokholyan NV, Shakhnovich EI, Shea J. 2005b. Reconstruction of the src-sh3 protein domain transition state ensemble using multiscale molecular dynamics simulations. J Mol Biol 350: 1035-1050.

Dobson CM. 1994. Protein folding. Solid evidence for molten globules. Curr Biol 4: 636-640.

Dokholyan NV. 2006. Studies of folding and misfolding using simplified models. Curr Opin Struct Biol 16: 79-85.

Duan Y, Kollman PA. 1998. Pathways to a protein folding intermediate observed in a 1-μs simulation in aqueous solution. Science 282: 740-744.

Dyson HJ, Wright PE. 1996. Insights into protein folding from NMR. Annu Rev Phys Chem 47: 369-395.

Dyson HJ, Wright PE. 2002. Coupling of folding and binding for unstructured proteins. Curr Opin Struct Biol 12: 54-60.

Ellgaard L, Ruddock LW. 2005. The human protein disulphide isomerase family: Substrate interactions and functional properties. EMBO Rep 6: 28-32.

Ellis RJ. 2005. Chaperone function: The orthodox view. Molecular Chaperones and Cell Signalling. Henderson B, Pockley AG, editors. Cambridge University Press; pp. 3-21.

Ellis RJ. 2006. Molecular chaperones: Assisting assembly in addition to folding. Trends Biochem Sci 31: 395-401.

Englander SW. 2000. Protein folding intermediates and pathways studied by hydrogen exchange. Annu Rev Biophys Biomol Struct 29: 213-238.

Englander SW, Mayne L. 1992. Protein folding studied using hydrogen-exchange labeling and two-dimensional NMR. Annu Rev Biophys Biomol Struct 21: 243-265.

Erickson HP. 1997. Stretching single protein molecules: Titin is a weird spring. Science 276: 1090-1092.

Evans MS, Clarke TF, Clark PL. 2005. Conformations of cotranslational folding intermediates. Protein Pept Lett 12: 189-195.

Evans PA, Dobson CM, Kautz RA, Hatfull G, Fox RO. 1987. Proline isomerism in staphylococcal nuclease characterized by NMR and site-directed mutagenesis. Nature 329: 266-268.

Eyles SJ. 2001. Proline not the only culprit? Nat Struct Biol 8: 380-381.

Eyles SJ, Kaltashov IA. 2004. Methods to study protein dynamics and folding by mass spectrometry. Methods 34: 88-99.

Fabian H, Naumann D. 2004. Methods to study protein folding by stopped-flow FT-IR. Methods 34: 28-40.

Farr GW, Furtak K, Rowland MB, Ranson NA, Saibil HR, et al. 2000. Multivalent binding of nonnative substrate proteins by the chaperonin GroEL. Cell 100: 561-573.

Fersht AR. 1995. Optimization of rates of protein folding: The nucleation-condensation mechanism and its implications. Proc Natl Acad Sci USA 92: 10869-10873.

Fersht AR. 1997. Nucleation mechanisms in protein folding. Curr Opin Struct Biol 7: 3-9.

Fersht AR, Daggett V. 2002. Protein folding and unfolding at atomic resolution. Cell 108: 573-582.

Fersht AR, Sato S. 2004. Φ-value analysis and the nature of protein-folding transition states. Proc Natl Acad Sci USA 101: 7976-7981.

Fersht AR, Matouschek A, Serrano L. 1992. The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. J Mol Biol 224: 771-782.

Fitch CA, Whitten ST, Hilser VJ, Garcia-Moreno EB. 2006. Molecular mechanisms of pH-driven conformational transitions of proteins: Insights from continuum electrostatics calculations of acid unfolding. Proteins 63: 113-126.

Flaherty KM, DeLuca-Flaherty C, McKay DB. 1990. Three-dimensional structure of the ATPase fragment of a 70k heat-shock cognate protein. Nature 346: 623-628.

Flynn GC, Pohl J, Flocco MT, Rothman JE. 1991. Peptide-binding specificity of the molecular chaperone bip. Nature 353: 726-730.

Forge V, Wijesinha RT, Balbach J, Brew K, Robinson CV, et al. 1999. Rapid collapse and slow structural reorganisation during the refolding of bovine α-lactalbumin. J Mol Biol 288: 673-688.

Franks F. 1995. Protein destabilization at low temperatures. Adv Protein Chem 46: 105-139.

Freedman RB. 1989. Protein disulfide isomerase: Multiple roles in the modification of nascent secretory proteins. Cell 57: 1069-1072.

Frydman J. 2001. Folding of newly translated proteins in vivo: The role of molecular chaperones. Annu Rev Biochem 70: 603-647.

Fulton KF, Main ER, Daggett V, Jackson SE. 1999. Mapping the interactions present in the transition state for unfolding/folding of fkbp12. J Mol Biol 291: 445-461.

Galzitskaya OV, Finkelstein AV. 1999. A theoretical search for folding/unfolding nuclei in three-dimensional protein structures. Proc Natl Acad Sci USA 96: 11299-11304.

Garel JR, Dautry-Varsat A. 1980. Sequential folding of a bifunctional allosteric protein. Proc Natl Acad Sci USA 77: 3379-3383.

Gejyo F, Yamada T, Odani S, Nakagawa Y, Arakawa M, et al. 1985. A new form of amyloid protein associated with chronic hemodialysis was identified as β_2-microglobulin. Biochem Biophys Res Commun 129: 701-706.

Gibson QH, Milnes L. 1964. Apparatus for rapid and sensitive spectrophotometry. Biochem J 91: 161-171.

Gilbert HF. 1997. Protein disulfide isomerase and assisted protein folding. J Biol Chem 272: 29399-29402.

Gillespie B, Plaxco KW. 2004. Using protein folding rates to test protein folding theories. Annu Rev Biochem 73: 837-859.

Gloss LM, Simler BR, Matthews CR. 2001. Rough energy landscapes in protein folding: Dimeric *E. coli* trp repressor folds through three parallel channels. J Mol Biol 312: 1121-1134.

Go N. 1983. Theoretical studies of protein folding. Annu Rev Biophys Bioeng 12: 183-210.

Goldenberg DP. 1999. Finding the right fold. Nat Struct Biol 6: 987-990.

Gothel SF, Marahiel MA. 1999. Peptidyl-prolyl *cis-trans* isomerases, a superfamily of ubiquitous folding catalysts. Cell Mol Life Sci 55: 423-436.

Grantcharova V, Alm EJ, Baker D, Horwich AL. 2001. Mechanisms of protein folding. Curr Opin Struct Biol 11: 70-82.

Grantcharova VP, Riddle DS, Santiago JV, Baker D. 1998. Important role of hydrogen bonds in the structurally polarized transition state for folding of the src SH3 domain. Nat Struct Biol 5: 714-720.

Grathwohl C, Wuthrich K. 1981. NMR studies of the rates of proline *cis-trans* isomerization in oligopeptides. Biopolymers 20: 2623-2633.

Gross M, Hessefort S. 1996. Purification and characterization of a 66-kDa protein from rabbit reticulocyte lysate which promotes the recycling of hsp 70. J Biol Chem 271: 16833-16841.

Gruebele M. 2002. Protein folding: The free energy surface. Curr Opin Struct Biol 12: 161-168.

Gruebele M. 2005. Downhill protein folding: Evolution meets physics. C R Biol 328: 701-712.

Hamada D, Segawa S, Goto Y. 1996. Non-native α-helical intermediate in the refolding of β-lactoglobulin, a predominantly β-sheet protein. Nat Struct Biol 3: 868-873.

Harrison SC, Durbin R. 1985. Is there a single pathway for the folding of a polypeptide chain?. Proc Natl Acad Sci USA 82: 4028-4030.

Hartl FU. 1996. Molecular chaperones in cellular protein folding. Nature 381: 571-579.

Heidary DK, Gross LA, Roy M, Jennings PA. 1997. Evidence for an obligatory intermediate in the folding of interleukin-1β. Nat Struct Biol 4: 725-731.

Herning T, Yutani K, Taniyama Y, Kikuchi M. 1991. Effects of proline mutations on the unfolding and refolding of human lysozyme: The slow refolding kinetic phase does not result from proline *cis-trans* isomerization. Biochemistry 30: 9882-9891.

Herold M, Leistler B, Hage A, Luger K, Kirschner K. 1991. Autonomous folding and coenzyme binding of the excised pyridoxal 5′-phosphate binding domain of aspartate aminotransferase from *Escherichia coli*. Biochemistry 30: 3612-3620.

Hesterkamp T, Hauser S, Lutcke H, Bukau B. 1996. *Escherichia coli* trigger factor is a prolyl isomerase that associates with nascent polypeptide chains. Proc Natl Acad Sci USA 93: 4437-4441.

Hobart SA, Ilin S, Moriarty DF, Osuna R, Colon W. 2002. Equilibrium denaturation studies of the *Escherichia coli* factor for inversion stimulation: Implications for in vivo function. Protein Sci 11: 1671-1680.

Hoerner JK, Xiao H, Kaltashov IA. 2005. Structural and dynamic characteristics of a partially folded state of ubiquitin revealed by hydrogen exchange mass spectrometry. Biochemistry 44: 11286-11294.

Hohfeld J, Jentsch S. 1997. GrpE-like regulation of the hsc70 chaperone by the anti-apoptotic protein bag-1. EMBO J 16: 6209-6216.

Hoshino M, Katou H, Hagihara Y, Hasegawa K, Naiki H, et al. 2002. Mapping the core of the β_2-microglobulin amyloid fibril by h/d exchange. Nat Struct Biol 9: 332-336.

Houry WA, Frishman D, Eckerskorn C, Lottspeich F, Hartl FU. 1999. Identification of in vivo substrates of the chaperonin GroEL. Nature 402: 147-154.

Huang GS, Oas TG. 1995. Submillisecond folding of monomeric λ repressor. Proc Natl Acad Sci USA 92: 6878-6882.

Hubner IA, Edmonds KA, Shakhnovich EI. 2005. Nucleation and the transition state of the SH3 domain. J Mol Biol 349: 424-434.

Hubner IA, Shimada J, Shakhnovich EI. 2004. Commitment and nucleation in the protein G transition state. J Mol Biol 336: 745-761.

Hughson FM, Wright PE, Baldwin RL. 1990. Structural characterization of a partly folded apomyoglobin intermediate. Science 249: 1544-1548.

Ikeguchi M, Kuwajima K, Mitani M, Sugai S. 1986. Evidence for identity between the equilibrium unfolding intermediate and a transient folding intermediate: A comparative study of the folding reactions of α-lactalbumin and lysozyme. Biochemistry 25: 6965-6972.

Isenman DE, Lancet D, Pecht I. 1979. Folding pathways of immunoglobulin domains. The folding kinetics of the cγ3 domain of human IgG1. Biochemistry 18: 3327-3336.

Jaenicke R. 1987. Folding and association of proteins. Prog Biophys Mol Biol 49: 117-237.

Jaenicke R. 1991. Protein folding: Local structures, domains, subunits, and assemblies. Biochemistry 30: 3147-3161.

Jaenicke R. 1999. Stability and folding of domain proteins. Prog Biophys Mol Biol 71: 155-241.

Jaenicke R. 2000. Stability and stabilization of globular proteins in solution. J Biotechnol 79: 193-203.

Jaenicke R, Lilie H. 2000. Folding and association of oligomeric and multimeric proteins. Adv Protein Chem 53: 329-401.

Jahn TR, Radford SE. 2005. The yin and yang of protein folding. FEBS J 272: 5962-5970.

Jamin M, Baldwin RL. 1998. Two forms of the pH 4 folding intermediate of apomyoglobin. J Mol Biol 276: 491-504.

Janin J, Wodak SJ. 1983. Structural domains in proteins and their role in the dynamics of protein function. Prog Biophys Mol Biol 42: 21-78.

Jecht M, Tomschy A, Kirschner K, Jaenicke R. 1994. Autonomous folding of the excised coenzyme-binding domain of D-glyceraldehyde-3-phosphate dehydrogenase from *Thermotoga maritima*. Protein Sci 3: 411-418.

Jorgensen WL, Gao J. 1988. *Cis-trans* energy difference for the peptide bond in the gas phase and in aqueous solution. J Am Chem Soc 110: 4212-4216.

Kad NM, Myers SL, Smith DP, Smith DA, Radford SE, et al. 2003. Hierarchical assembly of β_2-microglobulin amyloid in vitro revealed by atomic force microscopy. J Mol Biol 330: 785-797.

Kaltashov IA, Mohimen A. 2005. Estimates of protein surface areas in solution by electrospray ionization mass spectrometry. Anal Chem 77: 5370-5379.

Kamatari YO, Kitahara R, Yamada H, Yokoyama S, Akasaka K. 2004. High-pressure NMR spectroscopy for characterizing folding intermediates and denatured states of proteins. Methods 34: 133-143.

Kamatari YO, Yamada H, Akasaka K, Jones JA, Dobson CM, et al. 2001. Response of native and denatured hen lysozyme to high pressure studied by $^{15}N/^{1}H$ NMR spectroscopy. Eur J Biochem 268: 1782-1793.

Kandror O, Sherman M, Moerschell R, Goldberg AL. 1997. Trigger factor associates with GroEL in vivo and promotes its binding to certain polypeptides. J Biol Chem 272: 1730-1734.

Kang YK, Choi HY. 2004. *Cis-trans* isomerization and puckering of proline residue. Biophys Chem 111: 135-142.

Karplus M, Weaver DL. 1976. Protein-folding dynamics. Nature 260: 404-406.

Karplus M, Weaver DL. 1994. Protein folding dynamics: The diffusion-collision model and experimental data. Protein Sci 3: 650-668.

Kataoka M, Goto Y. 1996. X-ray solution scattering studies of protein folding. Fold Des 1: R107-R114.

Kazmirski SL, Daggett V. 1998a. Simulations of the structural and dynamical properties of denatured proteins: The "molten coil" state of bovine pancreatic trypsin inhibitor. J Mol Biol 277: 487-506.

Kazmirski SL, Daggett V. 1998b. Non-native interactions in protein folding intermediates: Molecular dynamics simulations of hen lysozyme. J Mol Biol 284: 793-806.

Kellermayer MS, Smith SB, Granzier HL, Bustamante C. 1997. Folding-unfolding transitions in single titin molecules characterized with laser tweezers. Science 276: 1112-1116.

Kelley RF, Richards FM. 1987. Replacement of proline-76 with alanine eliminates the slowest kinetic phase in thioredoxin folding. Biochemistry 26: 6765-6774.

Kelly SM, Price NC. 2000. The use of circular dichroism in the investigation of protein structure and function. Curr Protein Pept Sci 1: 349-384.

Khare SD, Ding F, Gwanmesia KN, Dokholyan NV. 2005. Molecular origin of polyglutamine aggregation in neurodegenerative diseases. PLoS Comput Biol 1: 230-235.

Kiefhaber T, Grunert HP, Hahn U, Schmid FX. 1992. Folding of RNase t1 is decelerated by a specific tertiary contact in a folding intermediate. Proteins 12: 171-179.

Kiefhaber T, Quaas R, Hahn U, Schmid FX. 1990a. Folding of ribonuclease T1. 1. Existence of multiple unfolded states created by proline isomerization. Biochemistry 29: 3053-3061.

Kiefhaber T, Quaas R, Hahn U, Schmid FX. 1990b. Folding of ribonuclease T1. 2. Kinetic models for the folding and unfolding reactions. Biochemistry 29: 3061-3070.

Kihara D, Lu H, Kolinski A, Skolnick J. 2001. Touchstone: An ab initio protein structure prediction method that uses threading-based tertiary restraints. Proc Natl Acad Sci USA 98: 10125-10130.

Kimura T, Akiyama S, Uzawa T, Ishimori K, Morishima I, et al. 2005. Specifically collapsed intermediate in the early stage of the folding of ribonuclease A. J Mol Biol 350: 349-362.

Kitahara R, Yamada H, Akasaka K, Wright PE. 2002. High pressure NMR reveals that apomyoglobin is an equilibrium mixture from the native to the unfolded. J Mol Biol 320: 311-319.

Kitahara R, Yokoyama S, Akasaka K. 2005. NMR snapshots of a fluctuating protein structure: Ubiquitin at 30 bar-3 kbar. J Mol Biol 347: 277-285.

Konermann L, Simmons DA. 2003. Protein-folding kinetics and mechanisms studied by pulse-labeling and mass spectrometry. Mass Spectrom Rev 22: 1-26.

Konishi Y, Ooi T, Scheraga HA. 1982. Regeneration of ribonuclease A from the reduced protein. Rate-limiting steps. Biochemistry 21: 4734-4740.

Kubota H, Hynes G, Willison K. 1995. The chaperonin containing t-complex polypeptide 1 (tcp-1). Multi-subunit machinery assisting in protein folding and assembly in the eukaryotic cytosol. Eur J Biochem 230: 3-16.

Kumar S, Nussinov R. 2002. Close-range electrostatic interactions in proteins. Chembiochem 3: 604-617.

Kunugi S, Tanaka N. 2002. Cold denaturation of proteins under high pressure. Biochim Biophys Acta 1595: 329-344.

Kuwajima K. 1996. The molten globule state of α-lactalbumin. FASEB J 10: 102-109.

Kuwajima K, Hiraoka Y, Ikeguchi M, Sugai S. 1985. Comparison of the transient folding intermediates in lysozyme and α-lactalbumin. Biochemistry 24: 874-881.

Kuwata K, Li H, Yamada H, Batt CA, Goto Y, et al. 2001. High pressure NMR reveals a variety of fluctuating conformers in β-lactoglobulin. J Mol Biol 305: 1073-1083.

Kuwata K, Li H, Yamada H, Legname G, Prusiner SB, et al. 2002. Locally disordered conformer of the hamster prion protein: A crucial intermediate to Prpsc? Biochemistry 41: 12277-12283.

Kuwajima K, Nitta K, Yoneyama M, Sugai S. 1976. Three-state denaturation of α-lactalbumin by guanidine hydrochloride. J Mol Biol 106: 359-373.

Lakey JH, Raggett EM. 1998. Measuring protein-protein interactions. Curr Opin Struct Biol 8: 119-123.

Lassalle MW, Li H, Yamada H, Akasaka K, Redfield C. 2003. Pressure-induced unfolding of the molten globule of all-Ala α-lactalbumin. Protein Sci 12: 66-72.

Lee S, Tsai FTF. 2005. Molecular chaperones in protein quality control. J Biochem Mol Biol 38: 259-265.

Levinthal C. 1969. How to fold graciously. DeBrunner P, Tsibris JCM, Munck E, editors. Mössbauer spectroscopy in biological systems: Proceedings of a meeting held at Allerton house, Monticello, Illinois. University of Illinois Press; pp. 22–24.

Levy Y, Cho SS, Onuchic JN, Wolynes PG. 2005. A survey of flexible protein binding mechanisms and their transition states using native topology based energy landscapes. J Mol Biol 346: 1121-1145.

Levy Y, Wolynes PG, Onuchic JN. 2004. Protein topology determines binding mechanism. Proc Natl Acad Sci USA 101: 511-516.

Lewis VA, Hynes GM, Zheng D, Saibil H, Willison K. 1992. T-complex polypeptide-1 is a subunit of a heteromeric particle in the eukaryotic cytosol. Nature 358: 249-252.

Li WF, Zhou XX, Lu P. 2005. Structural features of thermozymes. Biotechnol Adv 23: 271-281.

Lindorff-Larsen K, Rogen P, Paci E, Vendruscolo M, Dobson CM. 2005. Protein folding and the organization of the protein topology universe. Trends Biochem Sci 30: 13-19.

Lindorff-Larsen K, Vendruscolo M, Paci E, Dobson CM. 2004. Transition states for protein folding have native topologies despite high structural variability. Nat Struct Mol Biol 11: 443-449.

Lipman EA, Schuler B, Bakajin O, Eaton WA. 2003. Single-molecule measurement of protein folding kinetics. Science 301: 1233-1235.

Liu X, Siegel DL, Fan P, Brodsky B, Baum J. 1996. Direct NMR measurement of folding kinetics of a trimeric peptide. Biochemistry 35: 4306-4313.

Lomas DA, Carrell RW. 2002. Serpinopathies and the conformational dementias. Nat Rev Genet 3: 759-768.

Luders J, Demand J, Hohfeld J. 2000. The ubiquitin-related bag-1 provides a link between the molecular chaperones hsc70/hsp70 and the proteasome. J Biol Chem 275: 4613-4617.

Lue LF, Kuo YM, Roher AE, Brachova L, Shen Y, et al. 1999. Soluble amyloid β peptide concentration as a predictor of synaptic change in Alzheimer's disease. Am J Pathol 155: 853-862.

Lyon CE, Suh E, Dobson CM, Hore PJ. 2002. Probing the exposure of tyrosine and tryptophan residues in partially folded proteins and folding intermediates by CIDNP pulse-labeling. J Am Chem Soc 124: 13018-13024.

Ma H, Gruebele M. 2005. Kinetics are probe-dependent during downhill folding of an engineered λ6-85 protein. Proc Natl Acad Sci USA 102: 2283-2287.

Makhatadze GI, Privalov PL. 1995. Energetics of protein structure. Adv Protein Chem 47: 307-425.

Maki K, Ikura T, Hayano T, Takahashi N, Kuwajima K. 1999. Effects of proline mutations on the folding of staphylococcal nuclease. Biochemistry 38: 2213-2223.

Mandel S, Grunblatt E, Riederer P, Amariglio N, Jacob-Hirsch J, et al. 2005. Gene expression profiling of sporadic Parkinson's disease substantia nigra pars compacta reveals impairment of ubiquitin-proteasome subunits, skp1a, aldehyde dehydrogenase, and chaperone hsc-70. Ann N Y Acad Sci 1053: 356-375.

Marti DN, Bjelic S, Lu M, Bosshard HR, Jelesarov I. 2004. Fast folding of the HIV-1 and SIV gp41 six-helix bundles. J Mol Biol 336: 1-8.

Mateu MG, Sanchez Del Pino MM, Fersht AR. 1999. Mechanism of folding and assembly of a small tetrameric protein domain from tumor suppressor p53. Nat Struct Biol 6: 191-198.

Matouschek A, Kellis JTJ, Serrano L, Bycroft M, Fersht AR. 1990. Transient folding intermediates characterized by protein engineering. Nature 346: 440-445.

Matouschek A, Otzen DE, Itzhaki LS, Jackson SE, Fersht AR. 1995. Movement of the position of the transition state in protein folding. Biochemistry 34: 13656-13662.

Matthews CR, Hurle MR. 1987. Mutant sequences as probes of protein folding mechanisms. Bioessays 6: 254-257.

Mazhul' VM, Zaitseva EM, Shavlovsky MM, Stepanenko OV, Kuznetsova IM, et al. 2003. Monitoring of actin unfolding by room temperature tryptophan phosphorescence. Biochemistry 42: 13551-13557.

McCarney ER, Kohn JE, Plaxco KW. 2005. Is there or isn't there? The case for (and against) residual structure in chemically denatured proteins. Crit Rev Biochem Mol Biol 40: 181-189.

McLean CA, Cherny RA, Fraser FW, Fuller SJ, Smith MJ, et al. 1999. Soluble pool of Aβ amyloid as a determinant of severity of neurodegeneration in Alzheimer's disease. Ann Neurol 46: 860-866.

Meersman F, Smeller L, Heremans K. 2006. Protein stability and dynamics in the pressure-temperature plane. Biochim Biophys Acta 1764: 346-354.

Mersol JV, Steel DG, Gafni A. 1993. Detection of intermediate protein conformations by room temperature tryptophan phosphorescence spectroscopy during denaturation of *Escherichia coli* alkaline phosphatase. Biophys Chem 48: 281-291.

Miranker A, Robinson CV, Radford SE, Dobson CM. 1996. Investigation of protein folding by mass spectrometry. FASEB J 10: 93-101.

Missiakas D, Betton JM, Chaffotte A, Minard P, Yon JM. 1992. Kinetic studies of the refolding of yeast phosphoglycerate kinase: Comparison with the isolated engineered domains. Protein Sci 1: 1485-1493.

Mohimen A, Dobo A, Hoerner JK, Kaltashov IA. 2003. A chemometric approach to detection and characterization of multiple protein conformers in solution using electrospray ionization mass spectrometry. Anal Chem 75: 4139-4147.

Mok KH, Hore PJ. 2004. Photo-CIDNP NMR methods for studying protein folding. Methods 34: 75-87.

Mok KH, Nagashima T, Day IJ, Jones JA, Jones CJV, et al. 2003. Rapid sample-mixing technique for transient NMR and photo-CIDNP spectroscopy: Applications to real-time protein folding. J Am Chem Soc 125: 12484-12492.

Monera OD, Shaw GS, Zhu BY, Sykes BD, Kay CM, et al. 1992. Role of interchain α-helical hydrophobic interactions in Ca^{2+} affinity, formation, and stability of a two-site domain in troponin c. Protein Sci 1: 945-955.

Muller K, Garel JR. 1984. Folding of aspartokinase-homoserine dehydrogenase I is dominated by tertiary interactions. Biochemistry 23: 655-660.

Munoz V, Eaton WA. 1999. A simple model for calculating the kinetics of protein folding from three-dimensional structures. Proc Natl Acad Sci USA 96: 11311-11316.

Myers SL, Thomson NH, Radford SE, Ashcroft AE. 2006a. Investigating the structural properties of amyloid-like fibrils formed in vitro from β_2-microglobulin using limited proteolysis and electrospray ionisation mass spectrometry. Rapid Commun Mass Spectrom 20: 1628-1636.

Myers SL, Jones S, Jahn TR, Morten IJ, Tennent GA, et al. 2006b. A systematic study of the effect of physiological factors on β_2-microglobulin amyloid formation at neutral pH. Biochemistry 45: 2311-2321.

Nelson ED, Onuchic JN. 1998. Proposed mechanism for stability of proteins to evolutionary mutations. Proc Natl Acad Sci USA 95: 10682-10686.

Nelson R, Eisenberg D. 2006. Recent atomic models of amyloid fibril structure. Curr Opin Struct Biol 16: 260-265.

Niraula TN, Konno T, Li H, Yamada H, Akasaka K, et al. 2004. Pressure-dissociable reversible assembly of intrinsically denatured lysozyme is a precursor for amyloid fibrils. Proc Natl Acad Sci USA 101: 4089-4093.

Novokhatny VV, Kudinov SA, Privalov PL. 1984. Domains in human plasminogen. J Mol Biol 179: 215-232.

Nozaka M, Kuwajima K, Nitta K, Sugai S. 1978. Detection and characterization of the intermediate on the folding pathway of human α-lactalbumin. Biochemistry 17: 3753-758.

Odefey C, Mayr LM, Schmid FX. 1995. Non-prolyl *cis-trans* peptide bond isomerization as a rate-determining step in protein unfolding and refolding. J Mol Biol 245: 69-78.

Ohgushi M, Wada A. 1983. "Molten-globule state": A compact form of globular proteins with mobile side-chains. FEBS Lett 164: 21-24.

Onuchic JN, Wolynes PG. 2004. Theory of protein folding. Curr Opin Struct Biol 14: 70-75.

Onuchic JN, Socci ND, Luthey-Schulten Z, Wolynes PG. 1996. Protein folding funnels: The nature of the transition state ensemble. Fold Des 1: 441-450.

Ostermann J, Horwich AL, Neupert W, Hartl FU. 1989. Protein folding in mitochondria requires complex formation with hsp60 and ATP hydrolysis. Nature 341: 125-130.

Osvath S, Gruebele M. 2003. Proline can have opposite effects on fast and slow protein folding phases. Biophys J 85: 1215-1222.

Osvath S, Herenyi L, Zavodszky P, Fidy J, Kohler G. 2006. Hierarchic finite level energy landscape model—to describe

AU4

the refolding kinetics of phosphoglycerate kinase. J Biol Chem.

Osvath S, Sabelko JJ, Gruebele M. 2003. Tuning the heterogeneous early folding dynamics of phosphoglycerate kinase. J Mol Biol 333: 187-199.

Otzen DE, Kristensen O, Oliveberg M. 2000. Designed protein tetramer zipped together with a hydrophobic Alzheimer homology: A structural clue to amyloid assembly. Proc Natl Acad Sci USA 97: 9907-9912.

Pace CN, Shirley BA, McNutt M, Gajiwala K. 1996. Forces contributing to the conformational stability of proteins. FASEB J 10: 75-83.

Paci E, Clarke J, Steward A, Vendruscolo M, Karplus M. 2003. Self-consistent determination of the transition state for protein folding: Application to a fibronectin type III domain. Proc Natl Acad Sci USA 100: 394-399.

Pande VS, Baker I, Chapman J, Elmer SP, Khaliq S, et al. 2003. Atomistic protein folding simulations on the submillisecond time scale using worldwide distributed computing. Biopolymers 68: 91-109.

Parrini C, Taddei N, Ramazzotti M, Degl'Innocenti D, Ramponi G, et al. 2005. Glycine residues appear to be evolutionarily conserved for their ability to inhibit aggregation. Structure (Camb) 13: 1143-1151.

Pearl LH, Prodromou C. 2006. Structure and mechanism of the hsp90 molecular chaperone machinery. Annu Rev Biochem 75: 271-294.

Pedersen JS, Christensen G, Otzen DE. 2004. Modulation of s6 fibrillation by unfolding rates and gatekeeper residues. J Mol Biol 341: 575-588.

Peters TJ, Davidson LK. 1982. The biosynthesis of rat serum albumin. In vivo studies on the formation of the disulfide bonds. J Biol Chem 257: 8847-8853.

Petkova AT, Leapman RD, Guo Z, Yau W, Mattson MP, et al. 2005. Self-propagating, molecular-level polymorphism in Alzheimer's β-amyloid fibrils. Science 307: 262-265.

Pfeil W. 1981. The problem of the stability globular proteins. Mol Cell Biochem 40: 3-28.

Plaxco KW, Simons KT, Baker D. 1998. Contact order, transition state placement and the refolding rates of single domain proteins. J Mol Biol 277: 985-994.

Plotkin S, Wang J, Wolynes P. 1996. Correlated energy landscape model for finite, random heteropolymers. Phys Rev E Stat Nonlin Soft Matter Phys 53: 6271-6296.

Pokarowski P, Kolinski A, Skolnick J. 2003. A minimal physically realistic protein-like lattice model: Designing an energy landscape that ensures all-or-none folding to a unique native state. Biophys J 84: 1518-1526.

Prabhu NV, Sharp KA. 2005. Heat capacity in proteins. Annu Rev Phys Chem 56: 521-548.

Privalov PL. 1996. Intermediate states in protein folding. J Mol Biol 258: 707-725.

Ptitsyn OB. 1992. The molten globule state. Protein folding. Creighton TE, editor. New York: W.H. Freeman & Co.; pp. 243-300.

Ptitsyn OB. 1995. Molten globule and protein folding. Adv Protein Chem 47: 83-229.

Puett D. 1973. The equilibrium unfolding parameters of horse and sperm whale myoglobin. Effects of guanidine hydrochloride, urea, and acid. J Biol Chem 248: 4623-4634.

Radford SE, Dobson CM, Evans PA. 1992. The folding of hen lysozyme involves partially structured intermediates and multiple pathways. Nature 358: 302-307.

Raffen R, Dieckman LJ, Szpunar M, Wunschl C, Pokkuluri PR, et al. 1999. Physicochemical consequences of amino acid variations that contribute to fibril formation by immunoglobulin light chains. Protein Sci 8: 509-517.

Ramachandran GN, Mitra AK. 1976. An explanation for the rare occurrence of *cis* peptide units in proteins and polypeptides. J Mol Biol 107: 85-92.

Ramboarina S, Redfield C. 2003. Structural characterisation of the human α-lactalbumin molten globule at high temperature. J Mol Biol 330: 1177-1188.

Raschke TM, Kho J, Marqusee S. 1999. Confirmation of the hierarchical folding of RNase H: A protein engineering study. Nat Struct Biol 6: 825-831.

Razvi A, Scholtz JM. 2006. Lessons in stability from thermophilic proteins. Protein Sci 15: 1569-1578.

Redfield C. 2004. Using nuclear magnetic resonance spectroscopy to study molten globule states of proteins. Methods 34: 121-132.

Refaee M, Tezuka T, Akasaka K, Williamson MP. 2003. Pressure-dependent changes in the solution structure of hen egg-white lysozyme. J Mol Biol 327: 857-865.

Richardson JS, Richardson DC. 2002. Natural β-sheet proteins use negative design to avoid edge-to-edge aggregation. Proc Natl Acad Sci USA 99: 2754-2759.

Rief M, Gautel M, Oesterhelt F, Fernandez JM, Gaub HE. 1997. Reversible unfolding of individual titin immunoglobulin domains by AFM. Science 276: 1109-1112.

Robinson CR, Sauer RT. 1996. Equilibrium stability and submillisecond refolding of a designed single-chain arc repressor. Biochemistry 35: 13878-13884.

Robson B, Pain RH. 1976. The mechanism of folding of globular proteins. Equilibria and kinetics of conformational transitions of penicillinase from *Staphylococcus aureus* involving a state of intermediate conformation. Biochem J 155: 331-344.

Roder H, Elove GA, Englander SW. 1988. Structural characterization of folding intermediates in cytochrome *c* by H-exchange labelling and proton NMR. Nature 335: 700-704.

Roder H, Elöve GA, Shastry MC. R. 2000. Early stages of protein folding. Mechanisms of Protein Folding. Pain RH, editor. Oxford: Oxford University Press; pp. 65-104.

Roder H, Shastry MR. 1999. Methods for exploring early events in protein folding. Curr Opin Struct Biol 9: 620-626.

Roder H, Maki K, Cheng H. 2006. Early events in protein folding explored by rapid mixing methods. Chem Rev 106: 1836-1861.

Roder H, Maki K, Cheng H, Shastry MCR. 2004. Rapid mixing methods for exploring the kinetics of protein folding. Methods 34: 15-27.

Ross CA, Pickart CM. 2004. The ubiquitin-proteasome pathway in Parkinson's disease and other neurodegenerative diseases. Trends Cell Biol 14: 703-711.

Rounsevell R, Forman JR, Clarke J. 2004. Atomic force microscopy: Mechanical unfolding of proteins. Methods 34: 100-111.

Rousseau F, Schymkowitz JWH, Itzhaki LS. 2003. The unfolding story of three-dimensional domain swapping. Structure 11: 243-251.

Rudiger S, Germeroth L, Schneider-Mergener J, Bukau B. 1997. Substrate specificity of the DNAk chaperone determined by screening cellulose-bound peptide libraries. EMBO J 16: 1501-1507.

Rudolph R, Siebendritt R, Nesslauer G, Sharma AK, Jaenicke R. 1990. Folding of an all-β protein: Independent domain folding in γ II-crystallin from calf eye lens. Proc Natl Acad Sci USA 87: 4625-4629.

Sabelko J, Ervin J, Gruebele M. 1999. Observation of strange kinetics in protein folding. Proc Natl Acad Sci USA 96: 6031-6036.

Sali A, Shakhnovich E, Karplus M. 1994. How does a protein fold? Nature 369: 248-251.

Scharnagl C, Reif M, Friedrich J. 2005. Stability of proteins: Temperature, pressure and the role of the solvent. Biochim Biophys Acta 1749: 187-213.

Schellman JA. 1997. Temperature, stability, and the hydrophobic interaction. Biophys J 73: 2960-2964.

Scherer G, Kramer ML, Schutkowski M, Reimer U, Fischer G. 1998. Barriers to rotation of secondary amide peptide bonds. J Am Chem Soc 120: 5568-5574.

Schmid FX, Mayr LM, Mucke M, Schonbrunner ER. 1993. Prolyl isomerases: Role in protein folding. Adv Protein Chem 44: 25-66.

Schonbrunner ER, Schmid FX. 1992. Peptidyl-prolyl *cis-trans* isomerase improves the efficiency of protein disulfide isomerase as a catalyst of protein folding. Proc Natl Acad Sci USA 89: 4510-4513.

Schwaller M, Wilkinson B, Gilbert HF. 2003. Reduction-reoxidation cycles contribute to catalysis of disulfide isomerization by protein-disulfide isomerase. J Biol Chem 278: 7154-7159.

Schwartz R, Istrail S, King J. 2001. Frequencies of amino acid strings in globular protein sequences indicate suppression of blocks of consecutive hydrophobic residues. Protein Sci 10: 1023-1031.

Serpell LC, Sunde M, Benson MD, Tennent GA, Pepys MB, et al. 2000. The protofilament substructure of amyloid fibrils. J Mol Biol 300: 1033-1039.

Serrano L, Matouschek A, Fersht AR. 1992. The folding of an enzyme. III. structure of the transition state for unfolding of barnase analysed by a protein engineering procedure. J Mol Biol 224: 805-818.

Sha B, Lee S, Cyr DM. 2000. The crystal structure of the peptide-binding fragment from the yeast Hsp40 protein sis1. Structure 8: 799-807.

Shakhnovich EI, Finkelstein AV. 1989. Theory of cooperative transitions in protein molecules. I. Why denaturation of globular protein is a first-order phase transition. Biopolymers 28: 1667-1680.

Sharma AK, Minke-Gogl V, Gohl P, Siebendritt R, Jaenicke R, et al. 1990. Limited proteolysis of γ II-crystallin from calf eye lens. Physicochemical studies on the N-terminal domain and the intact two-domain protein. Eur J Biochem 194: 603-609.

Shastry MC, Luck SD, Roder H. 1998. A continuous-flow capillary mixing method to monitor reactions on the microsecond time scale. Biophys J 74: 2714-2721.

Shaw PE. 2002. Peptidyl-prolyl isomerases: A new twist to transcription. EMBO Rep 3: 521-526.

Shoelson SE, Sivaraja M, Williams KP, Hu P, Schlessinger J, et al. 1993. Specific phosphopeptide binding regulates a conformational change in the PI-3-kinase SH2 domain associated with enzyme activation. EMBO J 12: 795-802.

Shoemaker BA, Portman JJ, Wolynes PG. 2000. Speeding molecular recognition by using the folding funnel: The fly-casting mechanism. Proc Natl Acad Sci USA 97: 8868-8873.

Shortle D. 1996. The denatured state (the other half of the folding equation) and its role in protein stability. FASEB J 10: 27-34.

Shtilerman M, Lorimer GH, Englander SW. 1999. Chaperonin function: Folding by forced unfolding. Science 284: 822-825.

Siegert R, Leroux MR, Scheufler C, Hartl FU, Moarefi I. 2000. Structure of the molecular chaperone prefoldin: Unique interaction of multiple coiled coil tentacles with unfolded proteins. Cell 103: 621-632.

Silva JL, Cordeiro Y, Foguel D. 2006. Protein folding and aggregation: Two sides of the same coin in the condensation of proteins revealed by pressure studies. Biochim Biophys Acta 1764: 443-451.

Smeller L. 2002. Pressure-temperature phase diagrams of biomolecules. Biochim Biophys Acta 1595: 11-29.

Smith AV, Hall CK. 2001. Assembly of a tetrameric α-helical bundle: Computer simulations on an intermediate-resolution protein model. Proteins 44: 376-391.

Smith LJ, Fiebig KM, Schwalbe H, Dobson CM. 1996. The concept of a random coil. Residual structure in peptides and denatured proteins. Fold Des 1: R95-106.

Snow CD, Qiu L, Du D, Gai F, Hagen SJ, et al. 2004. Trp zipper folding kinetics by molecular dynamics and temperature-jump spectroscopy. Proc Natl Acad Sci USA 101: 4077-4082.

Song J, Bai P, Luo L, Peng ZY. 1998. Contribution of individual residues to formation of the native-like tertiary topology in the α-lactalbumin molten globule. J Mol Biol 280: 167-174.

Sorin EJ, Pande VS. 2005. Exploring the helix-coil transition via all-atom equilibrium ensemble simulations. Biophys J 88: 2472-2493.

Stefani M, Dobson CM. 2003. Protein aggregation and aggregate toxicity: New insights into protein folding, misfolding diseases and biological evolution. J Mol Med 81: 678-699.

Stein RL. 1993. Mechanism of enzymatic and nonenzymatic prolyl *cis-trans* isomerization. Adv Protein Chem 44: 1-24.

Steward A, Adhya S, Clarke J. 2002. Sequence conservation in Ig-like domains: The role of highly conserved proline residues in the fibronectin type III superfamily. J Mol Biol 318: 935-940.

Stewart DE, Sarkar A, Wampler JE. 1990. Occurrence and role of *cis* peptide bonds in protein structures. J Mol Biol 214: 253-260.

Stine WBJ, Dahlgren KN, Krafft GA, LaDu MJ. 2003. In vitro characterization of conditions for amyloid-β peptide oligomerization and fibrillogenesis. J Biol Chem 278: 11612-11622.

Street TO, Bradley CM, Barrick D. 2005. An improved experimental system for determining small folding entropy changes resulting from proline to alanine substitutions. Protein Sci 14: 2429-2435.

Sunde M, Blake C. 1997. The structure of amyloid fibrils by electron microscopy and X-ray diffraction. Adv Protein Chem 50: 123-159.

Szilagyi A, Zavodszky P. 2000. Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: Results of a comprehensive survey. Structure 8: 493-504.

Takahashi S, Yeh SR, Das TK, Chan CK, Gottfried DS, et al. 1997. Folding of cytochrome *c* initiated by submillisecond mixing. Nat Struct Biol 4: 44-50.

Takayama S, Bimston DN, Matsuzawa S, Freeman BC, Aime-Sempe C, et al. 1997. Bag-1 modulates the chaperone activity of hsp70/hsc70. EMBO J 16: 4887-4896.

Teale JM, Benjamin DC. 1977. Antibody as immunological probe for studying refolding of bovine serum albumin. Refolding within each domain. J Biol Chem 252: 4521-4526.

Terada K, Mori M. 2000. Human dnaj homologs dj2 and dj3, and bag-1 are positive cochaperones of hsc70. J Biol Chem 275: 24728-24734.

Teter SA, Houry WA, Ang D, Tradler T, Rockabrand D, et al. 1999. Polypeptide flux through bacterial hsp70: Dnak cooperates with trigger factor in chaperoning nascent chains. Cell 97: 755-765.

Texter FL, Spencer DB, Rosenstein R, Matthews CR. 1992. Intramolecular catalysis of a proline isomerization reaction in the folding of dihydrofolate reductase. Biochemistry 31: 5687-5691.

Tirado-Rives J, Jorgensen WL. 1993. Molecular dynamics simulations of the unfolding of apomyoglobin in water. Biochemistry 32: 4175-4184.

Todd MJ, Lorimer GH, Thirumalai D. 1996. Chaperonin-facilitated protein folding: Optimization of rate and yield by an iterative annealing mechanism. Proc Natl Acad Sci USA 93: 4030-4035.

Tsai C, Maizel JVJ, Nussinov R. 2002. The hydrophobic effect: A new insight from cold denaturation and a two-state water structure. Crit Rev Biochem Mol Biol 37: 55-69.

Tsai J, Levitt M, Baker D. 1999. Hierarchy of structure loss in MD simulations of src SH3 domain unfolding. J Mol Biol 291: 215-225.

Tsuji T, Chrunyk BA, Chen X, Matthews CR. 1993. Mutagenic analysis of the interior packing of an α/β barrel protein. Effects on the stabilities and rates of interconversion of the native and partially folded forms of the α subunit of tryptophan synthase. Biochemistry 32: 5566-5575.

Tsunenaga M, Goto Y, Kawata Y, Hamaguchi K. 1987. Unfolding and refolding of a type κ immunoglobulin light chain and its variable and constant fragments. Biochemistry 26: 6044-6051.

Tycko R. 2006. Molecular structure of amyloid fibrils: Insights from solid-state NMR. Q Rev Biophys: 1–55.

Udgaonkar JB, Baldwin RL. 1988. NMR evidence for an early framework intermediate on the folding pathway of ribonuclease A. Nature 335: 694-699.

Uversky VN, Li J, Souillac P, Millett IS, Doniach S, et al. 2002. Biophysical properties of the synucleins and their propensities to fibrillate: Inhibition of α-synuclein assembly by β- and γ-synucleins. J Biol Chem 277: 11970-11978.

Uzawa T, Akiyama S, Kimura T, Takahashi S, Ishimori K, et al. 2004. Collapse and search dynamics of apomyoglobin folding revealed by submillisecond observations of α-helical content and compactness. Proc Natl Acad Sci USA 101: 1171-1176.

Vanhove M, Lejeune A, Pain RH. 1998. β-lactamases as models for protein-folding studies. Cell Mol Life Sci 54: 372-377.

Vaucheret H, Signon L, Le Bras G, Garel JR. 1987. Mechanism of renaturation of a large protein, aspartokinase-homoserine dehydrogenase. Biochemistry 26: 2785-2790.

Vendruscolo M, Zurdo J, MacPhee CE, Dobson CM. 2003. Protein folding and misfolding: A paradigm of self-assembly and regulation in complex biological systems. Philos Transact A Math Phys Eng Sci 361: 1205-1222.

Villegas V, Martinez JC, Aviles FX, Serrano L. 1998. Structure of the transition state in the folding process of human procarboxypeptidase α2 activation domain. J Mol Biol 283: 1027-1036.

Wagner G, Wuthrich K. 1982. Amide protein exchange and surface conformation of the basic pancreatic trypsin inhibitor in solution. Studies with two-dimensional nuclear magnetic resonance. J Mol Biol 160: 343-361.

Walkenhorst WF, Green SM, Roder H. 1997. Kinetic evidence for folding and unfolding intermediates in staphylococcal nuclease. Biochemistry 36: 5795-5805.

Walker KW, Gilbert HF. 1997. Scanning and escape during protein-disulfide isomerase-assisted protein folding. J Biol Chem 272: 8845-8848.

Wand AJ, Roder H, Englander SW. 1986. Two-dimensional ^{1}H NMR studies of cytochrome *c*: Hydrogen exchange in the N-terminal helix. Biochemistry 25: 1107-1114.

Wedemeyer WJ, Welker E, Narayan M, Scheraga HA. 2000. Disulfide bonds and protein folding. Biochemistry 39: 4207-4216.

Wedemeyer WJ, Welker E, Scheraga HA. 2002. Proline *cis-trans* isomerization and protein folding. Biochemistry 41: 14637-14644.

Welker E, Maki K, Shastry MCR, Juminaga D, Bhat R, et al. 2004. Ultrarapid mixing experiments shed new light on the characteristics of the initial conformational ensemble during the folding of ribonuclease A. Proc Natl Acad Sci USA 101: 17681-17686.

Westermark P, Benson MD, Buxbaum JN, Cohen AS, Frangione B, et al. 2005. Amyloid: Toward terminology clarification. Report from the nomenclature committee of the international society of amyloidosis. Amyloid 12: 1-4.

Wetlaufer DB. 1973. Nucleation, rapid folding, and globular intrachain regions in proteins. Proc Natl Acad Sci USA 70: 697-701.

Williams KP, Shoelson SE. 1993. Cooperative self-assembly of SH2 domain fragments restores phosphopeptide binding. Biochemistry 32: 11279-11284.

Wolynes PG. 2005. Energy landscapes and solved protein-folding problems. Philos Transact A Math Phys Eng Sci 363: 453-464; discussion 464–467.

Wolynes PG, Onuchic JN, Thirumalai D. 1995. Navigating the folding routes. Science 267: 1619-1620.

Wong KB, Clarke J, Bond CJ, Neira JL, Freund SM, et al. 2000. Towards a complete description of the structural and dynamic properties of the denatured state of barnase and the role of residual structure in folding. J Mol Biol 296: 1257-1282.

Wong KP, Tanford C. 1973. Denaturation of bovine carbonic anhydrase B by guanidine hydrochloride. A process involving separable sequential conformational transitions. J Biol Chem 248: 8518-8523.

Wood LC, White TB, Ramdas L, Nall BT. 1988. Replacement of a conserved proline eliminates the absorbance-detected slow folding phase of iso-2-cytochrome *c*. Biochemistry 27: 8562-8568.

Wu Y, Matthews CR. 2002. Parallel channels and rate-limiting steps in complex protein folding reactions: Prolyl isomerization and the α subunit of trp synthase, a TIM barrel protein. J Mol Biol 323: 309-325.

Yamamoto S, Hasegawa K, Yamaguchi I, Tsutsumi S, Kardos J, et al. 2004. Low concentrations of sodium dodecyl sulfate induce the extension of β_2-microglobulin-related amyloid fibrils at a neutral pH. Biochemistry 43: 11075-11082.

Yang WY, Gruebele M. 2003. Folding at the speed limit. Nature 423: 193-197.

Yang WY, Gruebele M. 2004a. Detection-dependent kinetics as a probe of folding landscape microstructure. J Am Chem Soc 126: 7758-7759.

Yang WY, Gruebele M. 2004b. Folding λ-repressor at its speed limit. Biophys J 87: 596-608.

Zarrine-Afsar A, Larson SM, Davidson AR. 2005. The family feud: Do proteins with similar structures fold via the same pathway? Curr Opin Struct Biol 15: 42-49.

Zeeb M, Balbach J. 2004. Protein folding studied by real-time NMR spectroscopy. Methods 34: 65-74.

Zhang Y, Skolnick J. 2004. Automated structure prediction of weakly homologous proteins on a genomic scale. Proc Natl Acad Sci USA 101: 7594-7599.

Zhu X, Zhao X, Burkholder WF, Gragerov A, Ogata CM, et al. 1996. Structural analysis of substrate binding by the molecular chaperone Dnak. Science 272: 1606-1614.

Zimmerman SB, Trach SO. 1991. Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of *Escherichia coli*. J Mol Biol 222: 599-620.

Zimmerman SS, Scheraga HA. 1976. Stability of *cis*, *trans*, and nonplanar peptide groups. Macromolecules 9: 408-416.

Author Query Form

Neural Protein Metabolism and Function
Chapter No.: 10

Query Refs.	Details Required	Author's response
AU1	Should ''native fold'' be changed to ''native state''?	
AU2	Can ''....numerous examples, including...'' be changed to ''....numerous proteins, examples of which include....''?	
AU3	Should ''Rubisco'' be changed to ''RuBisCO''?	
AU4	Please provide page range and vol. no. for Osvath et al. (2006).	
AU5	Please provide publisher location.	