Understanding protein folding via free-energy surfaces from theory and experiment

Aaron R. Dinner, Andrej Šali,
Lorna J. Smith, Christopher M. Dobson
and Martin Karplus

The ability of protein molecules to fold into their highly structured functional states is one of the most remarkable evolutionary achievements of biology. In recent years, our understanding of the way in which this complex self-assembly process takes place has increased dramatically. Much of the reason for this advance has been the development of energy surfaces (landscapes), which allow the folding reaction to be described and visualized in a meaningful manner. Analysis of these surfaces, derived from the constructive interplay between theory and experiment, has led to the development of a unified mechanism for folding and a recognition of the underlying factors that control the rates and products of the folding process.

FOLLOWING ITS SYNTHESIS on the ribosome, a protein must fold to be functional. Although the cellular environment contains many factors that are involved in this process[1], the 'code' for folding is contained in the sequence, and many proteins have been shown to refold from denatured states in a test tube in the absence of such factors. In the past few years, advances in experiment and theory have led to considerable progress in our understanding of protein folding. There are two related 'protein folding problems': one concerns the mechanism by which proteins fold to their native structure[2–4] and the other the way in which the sequence selects that structure[5]. In this article we focus on the mechanism of folding and discuss some of the concepts that have emerged recently to enhance our understanding of this complex process.

We find it helpful to compare and contrast our knowledge of protein folding with ideas that were developed 30 years ago to describe simple chemical reactions[6]. All reactions take place on potential energy surfaces, which represent the energy of interaction of the atoms involved as a function of their positions. From such surfaces, the trajectories along which molecules move from reactants to products can be calculated, and the transition states and significantly populated intermediates (if any) that exist along the reaction pathway(s) can be determined. Although protein folding, which can involve several thousand coordinates, is clearly much more complex than a simple chemical reaction, energy surfaces play an essential role in both cases. However, one consequence of the much greater conformational freedom of polypeptide chains is that entropy is more important in protein folding than in most small-molecule reactions. As a result, for a folding reaction it is essential to consider the free energy rather than just the potential energy.

The concepts that are emerging from the nature of the surfaces, often pictorially referred to as energy landscapes[2,3], have resulted in a new view of protein folding[7]. This view has stimulated alternative interpretations of existing experiments and has led to the design of new experimental strategies to probe the details of the folding process[6]. Of particular significance is the emergence of a unified mechanism for folding that serves as a framework for further theoretical and experimental investigations of the behavior of specific protein molecules.

The protein folding reaction

To study protein folding *in vitro*, the solution conditions are generally changed from ones that stabilize the unfolded state to ones that stabilize the native state (e.g. by rapid dilution of denaturant). The reaction thus starts with a denatured polypeptide chain, which, in its limiting form, can be described as a random coil with no persistent long-range interactions[8,9]. It consists of a statistical distribution of rapidly interconverting states with different local and global properties. Fig. 1a shows the nature of the ensemble of conformations associated with a random-coil polypeptide chain under denaturing conditions. As can be seen from the figure, there is a dominant contribution from relatively extended conformations, but some conformations that are nearly as compact as the native structure also occur, although the probability of being in the native state itself is negligibly small.

Although the distribution of conformations in the initial random-coil state approximates to that expected for a simple organic polymer on a global level, each amino acid is limited to certain regions of backbone dihedral angle space at normal temperatures (Fig. 1b). Even if it is assumed that each amino acid populates only two low-energy regions of the Ramachandran plot[10,11], there are still of the order of 2^{100} or ~10^{30} main-chain conformations for a small protein with 100 residues. During the folding process, the polypeptide chain becomes highly compact, and each amino acid adopts its native geometry. If it is assumed that interconversion between main-chain conformations can occur as rapidly as the laws of physics allow (at a rate of ~$10^{11}\,\text{s}^{-1}$), it would take of the order of 10^{11} years to search 10^{30} conformations. The fact that folding to

A.R. Dinner*, L.J. Smith, C.M. Dobson and **M. Karplus*** are at the Oxford Centre for Molecular Sciences, New Chemistry Laboratory, University of Oxford, South Parks Road, Oxford, UK OX1 3QT; and **A. Šali** is at The Rockefeller University, 1230 York Avenue, New York, NY 10021-6399, USA.
Emails: chris.dobson@chem.ox.ac.uk (C.M.D.) and marci@tammy.harvard.edu (M.K.)
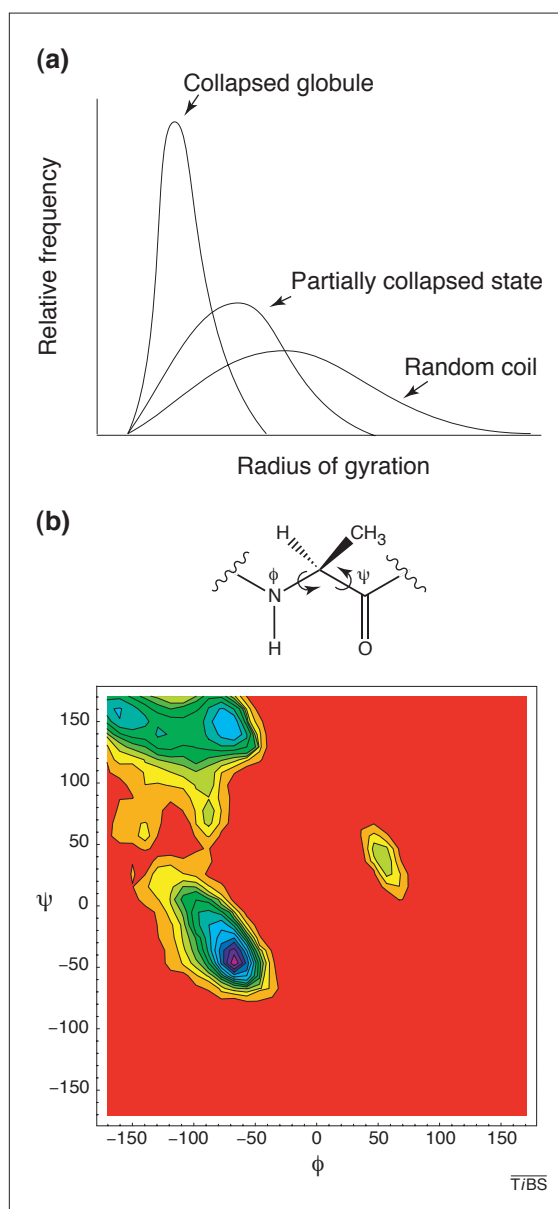* A.R. Dinner and M. Karplus are also at the Dept of Chemistry & Chemical Biology, Harvard University, 12 Oxford St, Cambridge, MA 02138, USA; and M. Karplus is also at the Laboratoire de Chimie Biophysique, Institut Le Bel, Université Louis Pasteur, 67000 Strasbourg, France.

the lowest-energy conformation occurs in a short time (typically in a second or less), in spite of the length of time needed for a systematic search of all conformations, has become known as the 'Levinthal Paradox'[12]. It is interesting to note that, with 10^{30} possible conformations, every molecule in a test tube, which contains at most 10^{18} molecules, could be in a different conformation in the random-coil state. Thus, describing the folding reaction requires a way of treating the behavior of a wide range of different structures. This poses significant challenges for both theory and experiment and, until recently, has limited progress in understanding protein folding.

Simulations of protein folding

It typically takes a millisecond or more for even a small protein to fold. However, with present-day computers, it is not possible to simulate the behavior of a protein for more than approximately one μs of physical time if the motions of all the atoms and the associated solvent molecules are considered explicitly[13]. It has been necessary therefore to develop simplified theoretical models to perform the large number of simulations required to obtain a meaningful description of the kinetics and thermodynamics of the folding reaction. Because of the simplifications, the results from such model calculations tend to be 'generic' in character. They suggest possible mechanisms of folding but cannot determine how a specific protein folds. For the latter, experiments and more detailed theoretical models are required, as we discuss below. Nevertheless, important insights have been obtained from simplified models.

One widely used class of models represents a protein as a string of beads positioned at sites on a lattice; at most, only one bead can occupy each site because of steric constraints. The beads interact with each other by pairwise contact potentials[2,3,6], which are effective energies (or 'potentials of mean force') in that they represent the residue–residue interactions in the presence of equilibrated solvent. The effective energies are usually chosen such that the beads attract each other, on

Figure 1
(a) Schematic representation of the global distribution of conformations of a polypeptide chain in a random-coil state, a partially collapsed denatured state and a compact non-native state. The different species within each ensemble interconvert rapidly. **(b)** Chemical structure of an alanine residue and the Ramachandran diagram representing its free-energy surface in a protein environment. The surface features result primarily from steric repulsions between the various atoms and lead to a distribution of conformations in the coil state that corresponds locally to an average over the low-energy regions of the conformational space for each amino acid. The diagram was obtained by taking the natural logarithm of the observed frequency of each pair of main-chain dihedral angles (ϕ,ψ) in a set of 1000 representative protein structures (A. Fiser, R.K.G. Do and A. Šali, unpublished). The contours are spaced 0.5 units apart, so that each level is a factor of $e^{0.5} = 1.6$ times more probable than the previous (lower) one.

average, which results in a lowest-energy conformation (the 'native state') that is fully compact (the $3 \times 3 \times 3$ cube at the bottom of Fig. 2), as an

approximation to the highly compact organized structure of a native protein.

The motions of the model polypeptide chain are simulated with a dynamic Monte Carlo algorithm[6]. Small random changes in the conformation of the chain are made repeatedly and accepted or rejected according to a rule, which is based on the change in energy. The algorithm is such that the probability is greater for the system to move to conformations of lower rather than higher energy. This mimics the situation in a real protein, where native-like interactions are on average more stabilizing than non-native ones. Many moves carried out in succession result in a folding trajectory that is directed by the potential energy function. The ability to 'watch' these models fold sets them apart from earlier phenomenological treatments in which molecular populations or concentrations of simplified folded fragments (e.g. helices) evolved according to diffusion equations[14]. The fact that the models do not involve specific assumptions about the folding process has been a major factor in the recent revolution in thinking about the mechanism of protein folding.
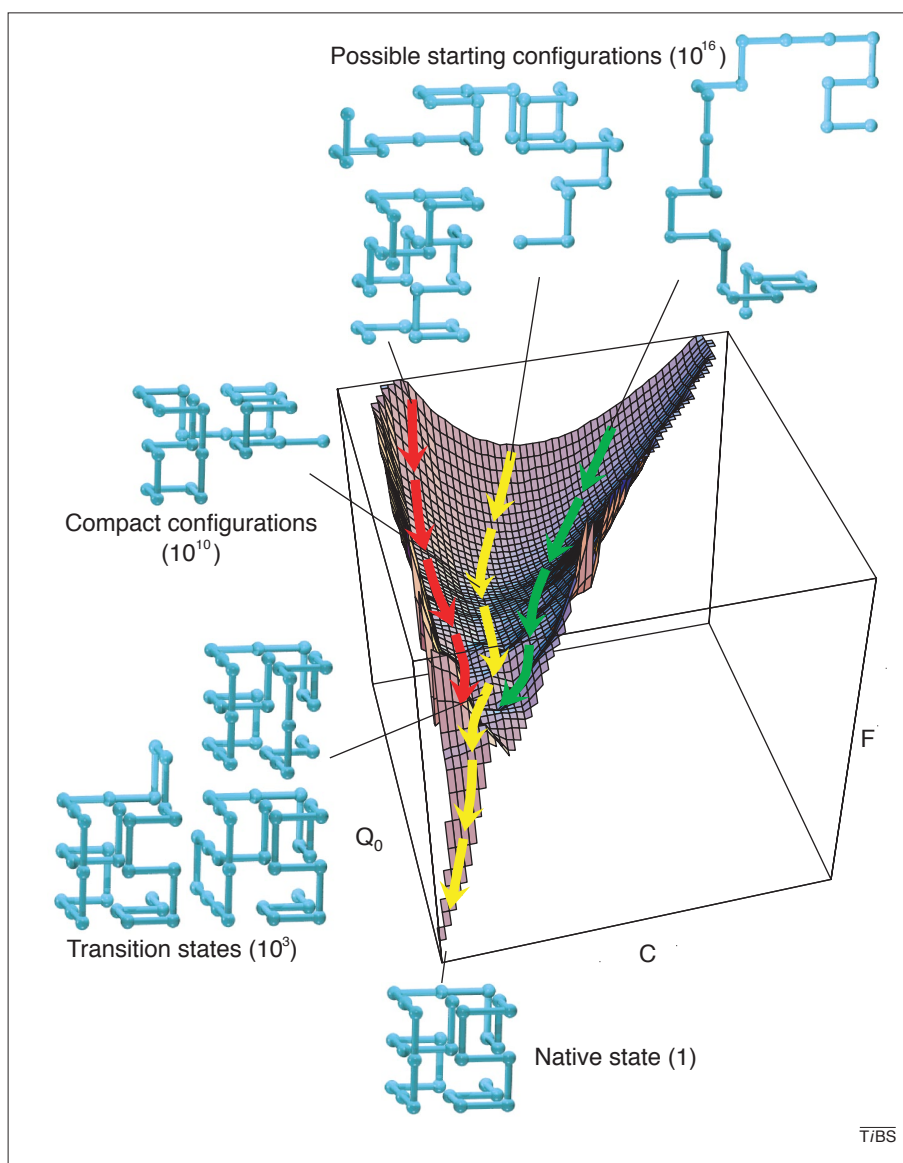
Because even simple molecules have many degrees of freedom, it is not feasible (or useful) to consider each conformation explicitly. Instead, it is necessary to introduce a reduced set of coordinates that characterizes the reaction and then to describe the full configuration space in terms of these coordinates. In the case of a simple chemical reaction, the choice of such coordinates is usually straightforward; for example, if the reaction involves the formation of a bond, a reasonable coordinate would be the distance between the two atoms involved. In the case of protein folding, the determination of the appropriate coordinate(s) is much more difficult[6,15]. A useful approach employs one or more 'progress variables', which describe the similarity of each conformation to the native state; such variables differ from true reaction coordinates in that they need not correspond to a minimum free-energy path through the transition state(s).

The free-energy surface calculated for the folding of a 27-mer from the simulations as a function of two commonly employed progress variables is shown in Fig. 2 (Ref. 16). The first coordinate, C, is the total number of contacts (native and non-native) between residues in the chain that are not covalently bonded to each other, and the second coordinate, Q_0, is the number of native contacts that are present in a given conformation. A typical folding simulation begins in a random conformation that is analogous to the state of a real protein molecule immediately following rapid dilution of denaturant. On average, the chain collapses rapidly to an ensemble of species with ~60% of the total number of possible contacts, of which only ~25% are nativelike (yellow trajectory in Fig. 2). At this stage of folding, there is a broad minimum on the free-energy surface associated with the ensemble of conformations that make up a disordered globule (Fig. 1a). The number of conformations readily accessible to the polypeptide chain is reduced from ~10^{16} to ~10^{10} by the collapse. After the collapse, the chain encounters the rate-limiting stage of the reaction, which is a random search through the semi-compact conformations for a transition state that leads rapidly to the native state.

Although in the case of small-molecule reactions the transition state is typically very restricted and geometrically well defined, this is not usually the case in protein folding. For the 27-mer, there are ~10^3 conformations that make up the transition-state ensemble, the common feature of which is that they comprise between 80 and 90% of the native contacts (three examples are shown on the left in Fig. 2). Given that the model polypeptide chain needs to find only one of the 10^3 transition states out of 10^{10} collapsed conformations, it can reach the native state in only $10^{10} / 10^3 = 10^7$ Monte Carlo moves, which is a small fraction of those that would be required to visit all the conformations (10^{16} or more). This means that the problem posed by the 'Levinthal Paradox' has been conceptually resolved: folding of the lattice model has demonstrated that it is possible to find the native state after searching through only a minute fraction of the total number of conformations[12].

Because the number of conformations at each successive stage of the folding process decreases (as indicated in Fig. 2), the entropy loss during folding plays an important role in determining the shape of the free-energy surface, in con-
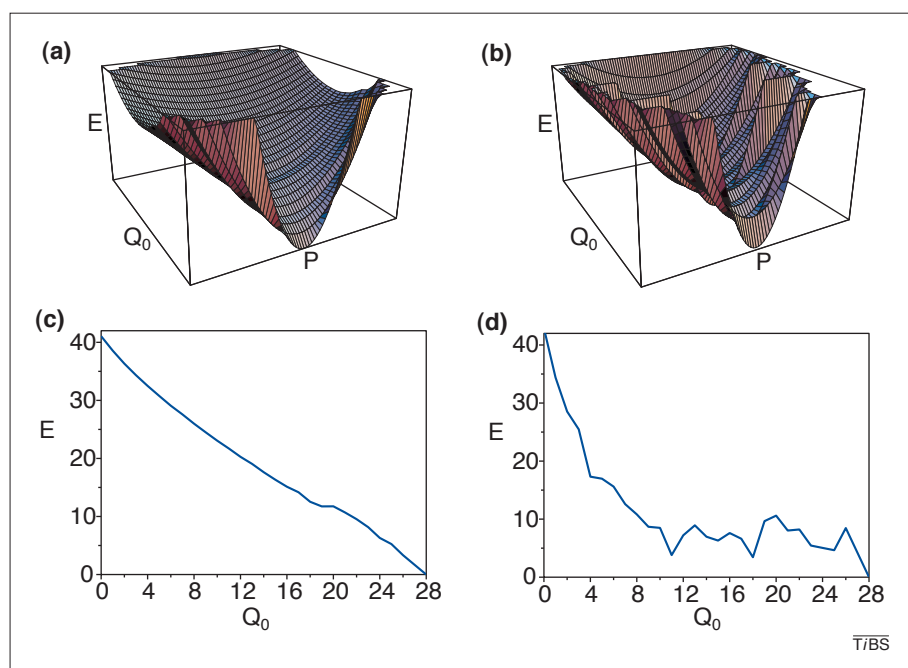

Figure 2
Free-energy (F) surface of a 27-mer as a function of the number of native contacts (Q_0) and the total number of (native and non-native) contacts (C) obtained by sampling the accessible configuration space with Monte Carlo simulations (see Ref. 16 for the method used). A fully extended chain has C = 0 (right-hand edge of surface), and a maximally compact 3 × 3 × 3 cube has C = 28 (left-hand edge of surface). The native state is a 3 × 3 × 3 cube (front left) with Q_0 = 28 (100%). The yellow trajectory shows the average path traced by the last structure sampled at each value of Q_0 [<C(Q_0)>] for 1000 independent trials that each began in a different random conformation. The other two trajectories (green and red) show a range of two standard deviations around the average and are thus expected to include ~95% of the trajectories. The structures illustrate the various stages of the reaction. From one of the 10^{16} possible random starting conformations, a folding chain collapses rapidly to a disordered globule. It then makes a slow, non-directed search among the 10^{10} semi-compact conformations for one of approximately 10^3 transition states that lead rapidly to the unique native state.

trast to most small-molecule reactions. To see how the energy and entropy act together to yield the minima and barriers in Fig. 2, it is useful to consider a diagram in which these components are shown separately[6] (Fig. 3). Two important cases are shown. In the first case (Fig. 3a), the effective energy decreases monotonically, on average, as the protein folds. We emphasize, on average, because each individual trajectory actually samples a series of microscopic local minima and barriers. If the contribution to the free energy of the configurational entropy decreases faster than the average energy, a 'bottleneck' results, and there is a barrier on the free-energy surface that corresponds to the transition states in Fig. 2. The form of the surface is such that the trajectory undergoes random motions that are biased towards the native state because, as pointed out

Figure 3

Quantitative schematic decomposition of the free energy of a 27-mer into its energetic and entropic components[6]. At each value of Q_0 (i.e. the number of native contacts), we draw a parabola whose lowest point corresponds to the average effective energy [$E(Q_0)$] and whose width depends on the configurational entropy [$S(Q_0)$]: $E(Q_0,P) = E(Q_0) + P^2[0.2/(S(Q_0) + 1)]$. P can be viewed as a measure of the width of the conformational ensemble for a given number of native contacts (Q_0) **(a)** The sequence and temperature are as in Fig. 2; the effective energy decreases steadily and the free-energy barrier to folding derives from the configurational entropy. **(b)** A sequence for which the native structure is the same but has been destabilized relative to the denatured state by increasing the dispersion of non-native interactions, so that non-native contacts can more readily compete with native ones; the folding time for this surface is determined primarily by the energetic cost for rearrangements between compact conformations. **(c)** Effective energy corresponding to (a); it shows a funnel-like behavior. **(d)** Effective energy corresponding to (b); it shows a rapid decrease followed by fluctuations resulting from interconversions between compact conformations.

above, native-like interactions are on average more stable than non-native ones. The second case (Fig. 3b) shows a surface from a simulation of a sequence for which the structure of the native state is the same as in Fig. 3a, but it has been destabilized relative to the denatured state. The effective energy decreases rapidly as the chain collapses to a disordered globule. After that, the energy remains relatively constant, with small barriers to further folding arising because non-native contacts (most of which stabilize a compact state) must be disrupted to allow native ones to be formed during the reorganization of the globule to give the native state. The barriers lead to a decrease in the rate of exploration of the energy surface[17]. A transition from a surface of the type shown in Fig. 3a to the type shown in Fig. 3b can be achieved for a given sequence simply by lowering the temperature[16].

The essence of the solution to the 'Levinthal Paradox' that has emerged from the lattice simulations is that an individual molecule needs to sample only a very small number of conformations because the nature of the effective energy surface restricts the search and there are many transition states. Individual polypeptide chains are likely to follow different trajectories in which the native contacts are formed in different orders. The trajectories tend to have more characteristics in common as the native state is approached. Because the energy and the number of conformations accessible to the ensemble of molecules decrease as the number of native contacts, Q_0, increases, as shown in Figs 3a and b, the term 'folding funnel'[18] has been introduced to describe the energy surface accessible to a polypeptide chain under refolding conditions. Whether the 'funneling' is significant (i.e. there is a faster reduction in the number of conformations that are sampled than is required from the native-like structure formed at each stage of the reaction) depends on the exact nature of the surface[6,19].

Experimental methods of studying folding

Experiments can be used to test the concepts arising from the simulations, but the methods of doing so are not straightforward because the population of polypeptide chains is extremely heterogeneous during the critical parts of the folding process. As a result, the conventional description of a 'structure' is often inadequate, and even the concept of an 'average structure' might be of limited value. Another complication is that, although folding is slow compared with many simple reactions, it is still fast (typically milliseconds to seconds) relative to the time scale of most structural methods applicable to macromolecules.

One approach to overcoming these problems is to use a variety of techniques to follow the development of different aspects of the formation of structure as a function of time. For example, far-UV circular dichroism (CD) can be used to determine the average content of secondary structure, whereas near-UV CD can be used to monitor the formation of close-packed structure around aromatic rings. Such methods have provided a global description with useful time resolution of the structural changes that take place during the folding reaction of several proteins[20]. Until recently, these approaches were limited to events on the millisecond timescale or longer, but now, methods using CD, fluorescence and infrared spectroscopy are being extended to study early (sub-millisecond) events in protein folding[21,22].

More detailed analysis of folding trajectories requires monitoring the nature and energies of the interactions between individual atoms as a function of time. There are two experimental approaches that can supply this type of information. One makes use of NMR spectroscopy and the other of site-directed mutagenesis (or protein engineering). The utility of the NMR method rests on its specificity at the level of individual atoms and on its ability, in principle, to determine the distribution of structures in a conformational ensemble (as well as the average structure) from parameters extracted from the spectra. The close proximity of pairs of atoms can be detected through the measurement of nuclear Overhauser effects (NOEs), and less specific information is available from measurement of, for example, hydrogen-exchange data, chemical shifts and relaxation phenomena. Limitations on the sensitivity and time resolution of such methods are being overcome, and this approach has great potential for the study of protein folding[23,24].

The other approach, the protein engineering method, depends on a

comparison of the properties of the wild-type protein with those of a series of mutants[25]. In brief, it is assumed that a residue is involved in native contacts in the transition state if a reduction in the size of its side chain by mutation (e.g. replacement of valine by alanine) destabilizes the transition state (as measured by a decrease in the folding rate) as much as it destabilizes the native state. Intermediates can be probed in an analogous manner, but the main strength of the method is its ability to shed light on the structure of the transition-state region[25]. A combination of the NMR and protein engineering approaches, complemented by the other methods already mentioned, is critical to linking experiments on specific proteins with theoretical descriptions of the folding reaction.

Linking theory and experiment

Simple folding reactions. From the co-incidence of the development of native-like character observed by different experimental techniques, the folding of many small proteins appears to be a fast, effectively two-state, reaction[26]. This means that the barrier to folding is low, and the population of protein molecules at any given stage of the reaction consists largely of either fully folded or essentially unfolded molecules. The free-energy surface in Fig. 2 is in accord with such simple two-state folding behavior. The calculated kinetics of folding on such a surface appear exponential for an ensemble of molecules, as do the observed kinetics of small simple and fast-folding proteins. Thus, in a global sense, the protein folding reaction can be deceptively similar in its kinetics to many small-molecule reactions.

The position of the rate-determining barrier in the folding reaction can vary with sequence. In the case of the 27-mer discussed above, the free-energy barrier that separates the folded and unfolded states occurs relatively late in the reaction. This is also true for many proteins, but in some cases, the transition state appears to involve only a small group of strong contacts (a 'nucleus'[25] or 'folding core'[27]), which simplifies the search process and increases the rate of the reaction. One protein for which such a nucleus is well documented is chymotrypsin inhibitor 2 (CI2)[28]. Although the important contacts are well defined, more realistic (all-atom) simulations of unfolding (rather than folding) indicate that only some of them are present in each member of the transition-state ensemble[29,30].

Even for a given sequence, the chain can be extended or collapsed when individual folding trajectories encounter the rate-determining barrier. In Fig. 2, collapse is fastest along the red trajectory, whereas it is slowest along the green one. The average behavior (yellow trajectory) depends on the relative strength of native versus non-native interactions. For a model of a three-helix bundle protein in which the residues were simplified but were not restricted to a lattice, it was found that the strength of the native contacts, relative to other contacts, determines whether the collapse results in a highly disordered or in a partially ordered (molten-globule-like) state (described in detail below)[31]. Whether or not the collapse precedes the formation of extensive native structure has important implications for the kinetics of folding. Experiments[32] and simulations[33] indicate that interconversion between conformations is slower in compact states than in more extended ones. Thus, folding along the red trajectory in Fig. 2 is expected to take longer on average than folding along the green one.
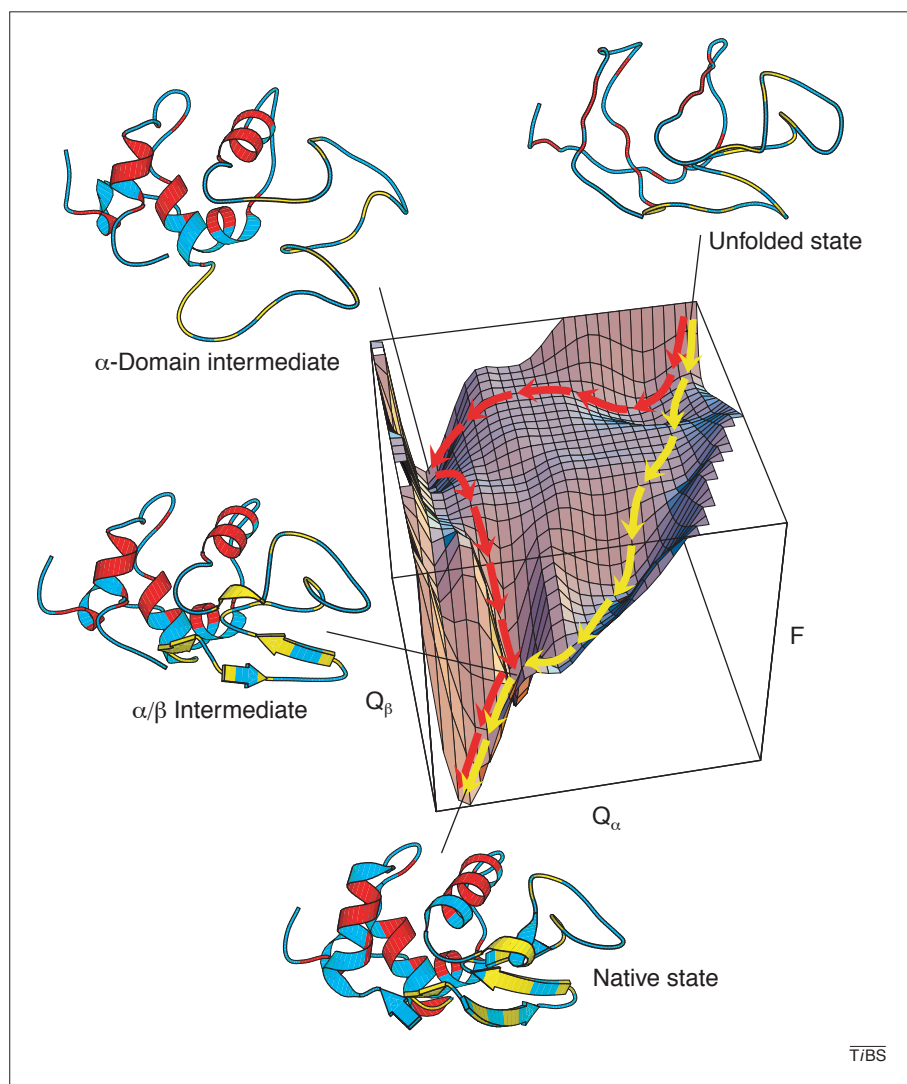
An experimental indicator of proteins that collapse rapidly to a partially organized compact state, which then converts more slowly to the native state (red trajectory in Fig. 2), is that the various spectroscopic probes exhibit different kinetics. Barstar appears to be a protein with such folding behavior[34]. Its folding corresponds to that expected for the energy surface shown in Fig. 3b. The roughness in the energy surface is associated with barriers to structural reorganization within the compact phase. Although included only implicitly in lattice models through the choice of pairwise interactions, one source of such barriers is the formation of contacts between side chains that are closely packed in the native state. In contrast to barstar, it has been shown with small-angle X-ray scattering (SAXS) that, in the IgG-binding domain of protein L, collapse and folding occur simultaneously (green trajectory in Fig. 2)[35].

The behavior of a given protein at physiological temperatures can be anywhere in the range of possibilities suggested by the two 'limiting' cases shown in Fig. 3. Moreover, mutations or changes in conditions that perturb the interactions between residues can result in a transition from a surface such as the one in Fig. 3b to one that corresponds to Fig. 3a. For example, mutation of a hydrophobic core residue in ubiquitin results in a shift from three-state to two-state kinetic behavior by destabilizing a partially folded intermediate[36].

Similarly, lowering the pH prevents a histidine in cytochrome *c* from forming a non-native ligation to the heme and thus accelerates folding to the native state[37]. More generally, it can be expected that, as more detailed microscopic information is obtained, the complex nature of the folding behavior of even such apparently simple fast-folding proteins will become apparent[38].

Complex folding reactions. As proteins increase in size, the barriers to reorganization in the collapsed state are likely to be larger, and this can introduce complexities in folding beyond those shown in Figs 2 and 3. The accessible regions of conformational space can become separated by high ridges, which results in multiple 'pathways' to the native state. In this case, the kinetics are likely to differ significantly from single exponential behavior, and distinct folding populations are possible due to the accumulation of long-lived intermediates along particular pathways[39]. A classic example of such a situation is the frequently observed heterogeneity in the folding rate of protein molecules associated with *cis–trans* isomerism of proline residues[40]. As in simple chemical reactions, intermediates often, but not always, tend to slow folding, although they can be important in ensuring that the final product (i.e. the protein native state) is formed in high yield[41].

An experimentally well-studied example of a protein that exhibits complexities in folding is provided by hen lysozyme, a small protein with two structural domains, one largely α helical (the α domain) and one with significant regions of β sheet (the β domain)[42]. Several phenomena are observed that are not seen in the folding of simpler proteins or predicted by the 27-mer lattice model discussed above. These phenomena include the presence of significantly populated intermediates in which a native-like structure is present in localized regions of the compact state. In addition, there is evidence for non-native-like interactions in these intermediates. Such interactions might be associated largely with interfacial regions that are not correctly formed in the absence of native-like contacts[43]. Moreover, the kinetics of lysozyme folding show extreme heterogeneity[42–44]. Approximately 20% of molecules fold to the native state with a time constant of less than 100 ms through a 'fast track' that includes a highly native-like intermediate[45]. Approximately 10% fold very slowly, in a

Figure 4

Schematic free-energy (F) surface representing features of the folding of hen lysozyme (a protein of 129 residues whose structure consists of two domains denoted α and β). $Q_\alpha(Q_\beta)$ is the number of native contacts in the α(β) domain. The yellow trajectory represents a 'fast track' in which the α and β domains form concurrently and populate the intermediate (labelled α/β) shown only briefly. The red trajectory represents a 'slow track' in which the chain becomes trapped in a long-lived intermediate with persistent structure in only the α domain; further folding from this intermediate involves either a transition over a higher barrier, or partial unfolding to enable the remainder of the folding process to occur along the fast track. Residues whose amide hydrogens are protected from solvent exchange in the native structure (as assessed by NMR) are colored red (α domain) or yellow (β domain); all others are blue. In each case, regions indicated to be native-like by monitoring the development of hydrogen exchange protection during kinetic refolding experiments[42–46] are drawn as ribbon representations of the native secondary structure elements (α helices and β sheets). The schematic structures of the various species were generated by constraining residues that are protected in those states to their conformations in the NMR structure[68] and heating the system to very high temperatures in molecular dynamics simulations. In the simulations, the disulfide bonds were kept intact, as they were in the experiments.

process that might be limited by proline isomerism in the manner discussed above. The majority of molecules (70%) populates an intermediate with persistent structure only in the α domain and then folds to the native state with a time constant of ~400 ms. Even here, however, kinetic heterogeneity exists as different populations of molecules form the intermediate at different rates[42]. This diversity in folding trajectories has been attributed to the disordered nature of the collapsed structure formed rapidly after folding is initiated and to the differences between the various intermediates that are formed during the folding process[42,43].

A schematic representation of a free-energy surface that explains at least qualitatively some of these features of the experimental data is shown in Fig. 4. Because lysozyme has at least two predominant pathways, at least two co-ordinates are needed to describe its folding behavior; we use the number of native-like contacts in each of the domains (Q_α and Q_β). The trajectories indicated on the energy surface illustrate two possible folding scenarios. Along the yellow trajectory, the α and β domains form concurrently, and the free energy decreases almost monotonically towards the native state (the 'fast track'). By contrast, along the red trajectory, formation of the α domain precedes that of the β domain, and the system becomes trapped in a free-energy minimum. Whether or not the α domain must partially unfold (indicated by a slight reversal in the red trajectory) to form the β domain (and, ultimately, the native state) remains unclear. The observed kinetic diversity discussed above suggests that a range of different events might be involved in overcoming the barriers to folding after formation of the stable α-domain intermediate. In all cases, the final step in folding appears to be the docking of the two domains to form the native structure[43–46].
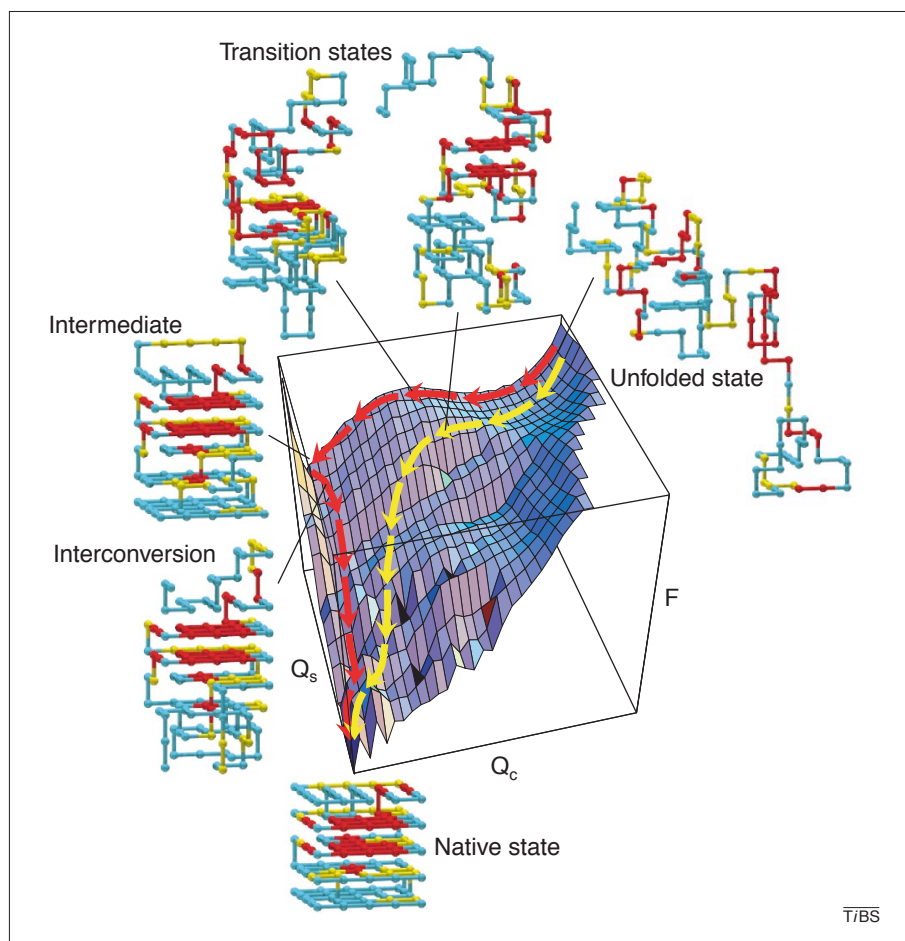
The experimentally established behavior for lysozyme (see Fig. 4) has a striking resemblance to independent simulation results for a lattice polymer model of comparable size: a 125-mer with a $5 \times 5 \times 5$ cubic native state[15,27]. As in the experimental example of lysozyme, the free-energy surface from the simulations is plotted as a function of two progress variables (Fig. 5), which were selected because they give a meaningful description of the folding reaction. The first of these is the number of 'core' (or 'nucleus') contacts (Q_c; residues shown in red), which were identified as those involving a minimal set of residues, which, upon formation of their native contacts, leads to consistent rapid folding. There are 34 such native contacts (out of a total of 176 in the fully compact state). The second progress variable is the number of native contacts in a set missing from the most prevalent folding intermediate (Q_s; residues shown in yellow). There are 15 such contacts; many, but not all, are on the surface.

Folding of the 125-mer begins with a rapid collapse to a semi-compact state followed by a slow, random search for the core; this part of the reaction is much like the complete folding of the 27-mer. Once ~60% of the core contacts are formed, the system reaches the transition states, which occur at a shallow entropic barrier on the free-energy

surface. Several major pathways for folding pass through this region. For a typical sequence, in ~15% of the trajectories, the chain finds several additional core contacts and folds directly to the native state (fast downhill folding along the yellow path). In another 40% of the trajectories, the chain forms particular native and non-native contacts outside the core (as illustrated by the transition state along the red trajectory), which divert the chain to a long-lived intermediate (an example of which is in the backleft corner of the surface in Fig. 5). In this intermediate, there are two subdomains (the first of which includes the three upper planes of the $5 \times 5 \times 5$ cube in Fig. 5, and the second of which includes the remainder of the residues) that are each substantially folded but are not packed against each other because of the presence of non-native structure. An energetic barrier must be overcome to escape from this state because stable contacts must be broken. In the remaining 45% of the trajectories, the chain also becomes partially misfolded but succeeds in escaping from the traps more rapidly. Statistical analysis of a database of 200 125-mer sequences has demonstrated the universality of the mechanism described above[27,47]. However, the specific fraction of fast- and slow-folding trajectories varies with sequence; because of the different roles played by the core and surface residues, the folding ability depends on both the secondary-structure content and the spatial distribution of low-energy contacts within the native state. This contrasts with results for the 27-mer, where no relationship was found between the folding rate and the secondary-structure content.

As suggested by the similarity of Figs 4 and 5, the folding mechanisms of the 125-mer lattice model and of lysozyme have many elements in common. In addition to the aspects already noted, in both cases, the solvent exposure of the transition state varies with temperature in a Hammond manner (i.e. the transition state moves towards the product as the latter is destabilized)[48,49], and the tendency to populate intermediates decreases as the temperature is raised[50]. Nevertheless, the relationship between the 125-mer and lysozyme is still at the level of an analogy; that is, the well-defined progress variables of the former need to be related more closely to the experimentally accessible variables of the latter. Additional experiments and simulations with more

Figure 5

Calculated free-energy surface for the folding of a 125-mer lattice model at a temperature equivalent to ~300 K. The free energy was obtained from a series of Monte Carlo simulations in which the protein was allowed to sample the full space of conformations[15]. It is plotted as a function of the number of 'core' contacts (denoted Q_c; residues shown in red) and the number of 'surface' contacts (denoted Q_s; residues shown in yellow); residues not involved in these variables are blue. The yellow trajectory represents an observed 'fast track' in which partial formation of the core leads directly to the native state, and the red trajectory represents a 'slow track' in which non-core contacts form before the core is complete and direct the chain to a misfolded intermediate that corresponds to a local free-energy minimum.

realistic models are in progress to achieve this objective.

Sequence, folding and structure

Determinants of structure. Determining how the native structure is encoded in the sequence of a protein is a major challenge, primarily because the specific fold is based on many weak (local) preferences for one structural feature over another. Understanding the folding code would provide insights into the way that evolution has given organisms the ability to synthesize a vast array of molecules with different functions by use of a rather subtle variation in the properties of the 'substituents' (the amino acid side chains) on an otherwise identical 'scaffold' (the polypeptide main chain). Moreover, it would allow us to exploit more effectively the immense body of data emerging from genome sequencing

projects[51] and, ultimately, to design proteins with novel properties. The mechanism of folding also provides insights as to how the structure is coded by the sequence.

In the early stages of folding, the weak preferences for different types of secondary structure in different regions of the sequence, coupled with the free-energy preference for burying hydrophobic residues and maintaining the solvent accessibility of polar residues, drive the protein molecules into native-like regions of conformational space. Such behavior has been seen in experiments[24,52] and in molecular dynamics simulations of peptides[53]. For proteins, the resulting state is believed to be similar in structure to 'equilibrium molten globules', which have a native-like topology but lack well-defined, unique tertiary interactions[54,55]. The

idea that the overall fold is determined by general aspects of the pattern of residues in the sequence, such as those enumerated above, is supported by the fact that it has been possible to design novel polypeptide sequences that fold into simple motifs (e.g. four helix bundles or β-hairpin structures)[56,57]. Such designed sequences usually (but not always[58]) yield molten globules rather than well-structured proteins.

Interconversion of a molten globule generated during folding to the fully native state requires the close packing of side chains. This final stage in the folding process tends to be highly cooperative and all residues are involved. An essential consequence of the balance between energy and entropy that leads to unique stable native states is that only one fold is of low enough energy for a given sequence. This requires that there be a sufficient difference in energy between the native state and the many structures that can exist without close packing (an 'energy gap')[2,16]. Alternative close-packed folds of natural proteins will not normally be able to accommodate all the residues without large energy penalties, which eliminates the possibility of misfolding into very different, but well-defined, structures.

Determinants of folding rate. For small two-state proteins, it has been found that the folding rate is correlated with the contact order (the average separation in sequence of atomic contacts)[59]. In general, α-helical proteins, which have a larger fraction of sequentially short-range contacts, fold faster than mixed-α/β or β-sheet proteins, which have a larger fraction of long-range contacts. Although lattice models do not fully capture the behavior of α helices, the first study to show statistically that short-range cooperative contacts accelerate folding was of the 125-residue lattice model described above[27,47]. The relationship between kinetic behavior and equilibrium structure derives from the fact that the transition state involves many of the native interactions, as observed in both experiments[25] and simulations[15], so that the transition state has a topology close to that of the native state. For a folding reaction that is rate-limited by an entropic barrier[15], increasing the fraction of contacts that are short-range in the native state (and thus the transition state) accelerates folding in two related ways: (1) it facilitates locating the transition state by a diffusive search, and (2) it lowers the entropic barrier to folding[60]. It is important to

note, however, that the absence of non-native interactions in the transition state is not inconsistent with their presence in off-pathway intermediates that lead to more complex kinetics[15].

In the excitement surrounding the simple empirical relationship between folding rate and contact order, stability is often neglected[59]. However, point mutations are known to have significant effects on the folding rate because the interactions involved in the folding core or nucleus depend on specific features of the side chains. Indeed, this is the basis for protein engineering experiments, which show that folding rates exhibit an approximately exponential dependence on changes in free energy[60]. The important role of stability in determining kinetic behavior, which was observed for lattice models, has been verified by incorporation of the free energy of unfolding into improved models for fitting experimental rates[61]. In any case, the relatively simple dependence of the folding rate on a small number of native properties is in accord with the unified mechanism for protein folding discussed above.

For small proteins, there is usually a single folding nucleus, and the degree of native-like structure is likely to be high enough to give rise to rapid subsequent folding and hence to two-state behavior. For larger proteins, multiple nucleation sites are likely to exist and give rise to more complex behavior, including a range of intermediates[62]. In either case, structural fluctuations must still be large enough to enable the molecules to explore different conformations until the lowest-energy state is reached. For larger proteins with multiple nucleation sites, this step is often the slowest one in the entire folding process.

Outlook

Schematic and simulated free-energy surfaces, such as those in Figs 4 and 5, allow the new view of folding to be compared with the more conventional picture. The concepts used in the latter, such as pathways, transition states, domains and intermediates, all remain meaningful in the new view, albeit in an ensemble context. Phenomenological models remain useful for describing the folding mechanisms of specific proteins and for interpreting experiments. For example, it has been shown recently that the diffusion collision model[14] gives a quantitative explanation of the large changes in rate observed for certain mutations of monomeric λ repressor[63]. An

essential point for the validity of such a description, in accord with the new view, is that the energy decreases as the various structural elements are formed in such a way as to guide the system to the native state.

If the shape of the free-energy surface for protein folding is such that the trajectories of all the molecules are limited to a narrow region of conformational space, the description of folding in terms of a single pathway is likely to be satisfactory. However, even with relatively well-defined pathways, the excursions are expected to be much larger than those from the minimum free-energy pathway in small-molecule reactions. For surfaces on which the trajectories of different molecules pass through very broad regions of conformational space, a simple pathway description of folding might be inappropriate and could lead to incorrect interpretation of experimental data. An essential aspect of the new view of folding is that the different types of behavior observed in a variety of experimental studies are a consequence of the relative importance of different contributions to the free energy, not of fundamentally different mechanisms. The new view thus serves to unify the field in terms of a general mechanism of protein folding.

Bringing together the recent advances in theory and experiment through the concepts of energy and free energy surfaces has provided an understanding of the essential elements of the mechanism of protein folding and has led to a framework for future research. It is becoming increasingly evident that not only is folding important for generating functional proteins, but, along with unfolding, it is an integral part of a range of biological processes ranging from molecular trafficking to the control of the cell cycle[1,64]. Moreover, it is now well established that a range of diseases, from cystic fibrosis to Alzheimer's and from cancer to the transmissible spongiform encephalophathies, is associated with the failure of protein folding and unfolding to be controlled in an appropriate manner[64–66]. An increased knowledge of these complex processes at a molecular level[67] is therefore expected to lead to fundamental advances in our understanding of many aspects of biology and medicine.

References

1 Ellis, R.J. and Hartl, F.U. (1999) Principles of protein folding in the cellular environment. *Curr. Opin. Struct. Biol.* 9, 102–110

2 Bryngelson, J.D. *et al.* (1995) Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins* 21, 167–195

3 Dill, K.A. and Chan, H.S. (1997) From Levinthal to pathways to funnels. *Nature Struct. Biol.* 4, 10–19

4 Dobson, C.M. and Karplus, M. (1999) The fundamentals of protein folding: bringing together theory and experiment. *Curr. Opin. Struct. Biol.* 9, 92–101

5 Sippl, M.J. *et al.* (1999) An attempt to analyse progress in fold recognition from CASP1 to CASP3. *Proteins* Suppl. 3, 226–230

6 Dobson, C.M. *et al.* (1998) Protein folding: a perspective from theory and experiment. *Angew. Chem. Int. Ed. Eng.* 37, 868–893

7 Baldwin, R.L. (1994) Protein folding. Matching speed and stability. *Nature*, 369, 183–184

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

9 Smith, L.J. *et al.* (1996) The concept of a random coil. Residual structure in peptides and denatured proteins. *Fold. Des.* 1, R95–R106

10 Daura, X. *et al.* (1998) Reversible peptide folding in solution by molecular dynamics simulation. *J. Mol. Biol.* 280, 925–932

11 Dinner, A.R. *et al.* (1999) Understanding β-hairpin formation. *Proc. Natl. Acad. Sci. U. S. A.* 96, 9068–9073

12 Karplus, M. (1997) The Levinthal paradox: yesterday and today. *Fold. Des.* 2, 569–576

13 Duan, Y. and Kollman, P.A. (1998) Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science* 282, 740–744

14 Karplus, M. and Weaver, D.L. (1976). Protein-folding dynamics. *Nature* 260, 404–406

15 Dinner, A.R. and Karplus, M. (1999) The thermodynamics and kinetics of protein folding: a lattice model analysis of multiple pathways with intermediates. *J. Phys. Chem. B* 103, 7976–7994

16 Šali, A. *et al.* (1994) How does a protein fold? *Nature* 369, 248–251.

17 Socci, N.D. *et al.* (1996) Diffusive dynamics of the reaction coordinate for protein folding. *J. Chem. Phys.* 104, 5860–5868

18 Wolynes, P.G. *et al.* (1995) Navigating the folding routes. *Science* 267, 1619–1620

19 Finkelstein, A.V. and Badretdinov, A.Y. (1997) Rate of protein folding near the point of thermodynamic equilibrium between the coil and the most stable chain fold. *Fold. Des.* 2, 115–121

20 Plaxco, K.W. and Dobson, C.M. (1996) Time-resolved biophysical methods in the study of protein folding. *Curr. Opin. Struct. Biol.* 6, 630–636

21 Eaton, W.A. *et al.* (1997) Submillisecond kinetics of protein folding. *Curr. Opin. Struct. Biol.* 7, 10–14

22 Callender, R.H. *et al.* (1998) Fast events in protein folding: the time evolution of primary processes. *Annu. Rev. Phys. Chem.* 49, 173–202

23 Dyson, H.J. and Wright, P.E. (1996) Insights into protein folding from NMR. *Annu. Rev. Phys. Chem.* 47, 369–395

24 Dobson, C.M. and Hore, P.J. (1998) Kinetic studies of protein folding using NMR spectroscopy. *Nature Struct. Biol.* 5, 504–507

25 Fersht, A. (1999) *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding.* W. H. Freeman & Co., New York

26 Jackson, S. E. (1998) How do small single-domain proteins fold? *Fold. Des.* 3, R81–R91

27 Dinner, A.R. *et al.* (1996). The folding mechanism of larger model proteins: role of native structure. *Proc. Natl. Acad. Sci. U. S. A.* 93, 8356–8361

28 Itzhaki, L.S. *et al.* (1995) The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: evidence for a nucleation-condensation mechanism for protein folding. *J. Mol. Biol.* 254, 260–288

29 Lazaridis, T. and Karplus, M. (1997) 'New view' of protein folding reconciled with the old through multiple unfolding simulations. *Science* 278, 1928–1931

30 Daggett, V. *et al.* (1996) Structure of the transition state for folding of a protein derived from experiment and simulation. *J. Mol. Biol.* 257, 430–440

31 Zhou, Y. and Karplus, M. (1999) Interpreting the folding kinetics of helical proteins. *Nature* 401, 400–403

32 Baum, J. *et al.* (1989) Characterization of a partly folded protein by NMR methods: studies on the molten globule state of guinea pig α-lactalbumin. *Biochemistry*, 28, 7–13

33 Gutin, A.M. *et al.* (1995) Is burst hydrophobic collapse necessary for protein folding? *Biochemistry* 34, 3066–3076

34 Agashe, V.R. *et al.* (1995) Initial hydrophobic collapse in the folding of barstar. *Nature* 377, 754–757

35 Plaxco, K.W. *et al.* (1999) Chain collapse can occur concomitantly with the rate-limiting step in protein folding. *Nature Struct. Biol.* 6, 554–556

36 Khorasanizadeh, S. *et al.* (1996) Evidence for a three-state model of protein folding from kinetic analysis of ubiquitin variants with altered core residues. *Nature Struct. Biol.* 3, 193–205

37 Sosnick, T.R. *et al.* (1994) The barriers in protein folding. *Nature Struct. Biol.* 1, 149–156

38 Sabelko, J. *et al.* (1999) Observation of strange kinetics in protein folding. *Proc. Natl. Acad. Sci. U. S. A.* 96, 6031–6036

39 Kim, P.S. and Baldwin, R.L. (1990) Intermediates in the folding reactions of small proteins. *Annu. Rev. Biochem.* 59, 631–660

40 Schmid, F.X. (1992) Kinetics of unfolding and refolding of single-domain proteins. In *Protein Folding* (T. E. Creighton, ed.), pp. 197–241, W. H. Freeman & Co., New York

41 Ferguson, N. *et al.* (1999) Rapid folding with and without populated intermediates in the homologous four-helix proteins Im7 and Im9. *J. Mol. Biol.* 286, 1597–1608

42 Radford, S.E. *et al.* (1992) The folding of hen lysozyme involves partially structured intermediates and multiple pathways. *Nature* 358, 302–307

43 Matagne, A. *et al.* (1997) Fast and slow tracks in lysozyme folding: insight into the role of domains in the folding process. *J. Mol. Biol.* 267, 1068–1074

44 Kiefhaber, T. (1995) Kinetic traps in lysozyme folding. *Proc. Natl. Acad. Sci. U. S. A.* 92, 9029–9033

45 Kulkarni, S.K. *et al.* (1999) A near-native state on the slow refolding pathway of hen lysozyme. *Prot. Sci.* 8, 35–44

46 Segel, D.J. *et al.* (1999) Characterization of transient intermediates in lysozyme folding with time-resolved small-angle X-ray scattering. *J. Mol. Biol.* 288, 489–499

47 Dinner, A.R. *et al.* (1998) Use of quantitative structure-property relationships to predict the folding ability of model proteins. *Proteins* 33, 177–203

48 Dinner, A.R. and Karplus, M. (1999) Is protein unfolding the reverse of protein folding? A lattice simulation analysis. *J. Mol. Biol.* 292, 403–419

49 Ibarra-Molero, B. and Sanchez-Ruiz, J.M. (1996) A model-independent, nonlinear extrapolation procedure for the characterization of protein folding energetics from solvent-denaturation data. *Biochemistry* 35, 14689–14702

50 Matagne, A. *et al.* (2000) Thermal unfolding of an intermediate is associated with non-Arrhenius kinetics in the folding of hen lysozyme. *J. Mol. Biol.* 297, 193–210

51 Burley, S.K. *et al.* (1999) Structural genomics: beyond the human genome project. *Nat. Genet.* 23, 151–157

52 Dyson, H.J. and Wright, P.E. (1998) Equilibrium NMR studies of unfolded and partially folded proteins. *Nature Struct. Biol.* 5, 499–503

53 Schaefer, M. *et al.* (1998) Solution conformations and thermodynamics of structured peptides: molecular dynamics simulation with an implicit solvation model. *J. Mol. Biol.* 284, 835–847

54 Ptitsyn, O.B. (1995) Structures of folding intermediates. *Curr. Opin. Struct. Biol.* 5, 74–78

55 Kuwajima, K. (1996) The molten globule state of α-lactalbumin. *FASEB J.* 10, 102–109

56 Kamtekar, S. *et al.* (1993) Protein design by binary patterning of polar and nonpolar amino acids. *Science* 262, 1680–1685

57 Hecht, M.H. *et al.* (1990) De novo design, expression, and characterization of Felix: a four-helix bundle protein of native-like sequence. *Science* 249, 884–891

58 Roy, S. *et al.* (1997) A protein designed by binary patterning of polar and nonpolar amino acids displays native-like properties. *J. Am. Chem. Soc.* 119, 5302–5306

59 Plaxco, K.W. *et al.* (1998) Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol.* 277, 985–994

60 Fersht, A.R. (2000) Transition-state structure as a unifying basis in protein-folding mechanisms: contact order, chain topology, stability, and the extended nucleus mechanism. *Proc. Natl. Acad. Sci. U. S. A.* 97, 1525–1529

61 Muñoz, V. and Eaton, W.A. (1999) A simple model for calculating the kinetics of protein folding from three-dimensional structures. *Proc. Natl. Acad. Sci. U. S. A.* 96, 11311–11316

62 Morozova-Roche, L.A. *et al.* (1999) Independent nucleation and heterogeneous assembly of structure during folding of equine lysozyme. *J. Mol. Biol.* 289, 1055–1073

63 Burton, R.E. *et al.* (1998) Protein folding dynamics: quantitative comparison between theory and experiment. *Biochemistry* 37, 5337–5343

64 Radford, S.E. and Dobson, C.M. (1999) From computer simulations to human disease: emerging themes in protein folding. *Cell* 97, 291–298

65 Thomas, P.J. *et al.* (1995) Defective protein folding as a basis of human disease. *Trends Biochem. Sci.* 20, 456–459

66 Dobson, C.M. (1999) Protein misfolding, evolution and disease. *Trends Biochem. Sci.* 24, 329–332

67 Dinner, A.R. and Karplus, M. (1998) A metastable state in folding simulations of a protein model. *Nature Struct. Biol.* 5, 236–241

68 Smith, L.J. *et al.* (1993) Structure of hen lysozyme in solution. *J. Mol. Biol.* 229, 930–944