Theoretical studies of protein folding and unfolding

Martin Karplus and Andrej Šali

Université Louis Pasteur, Strasbourg, France and The Rockefeller University, New York, USA

The mechanism of protein folding is being investigated theoretically by the use of both simplified and all-atom models of the polypeptide chain. Lattice heteropolymer simulations of the folding process have led to proposals for the folding mechanism and for the resolution of the Levinthal paradox. Both stability and rapid folding have been shown in model studies to result from the presence of a pronounced global energy minimum corresponding to the native state. Concomitantly, molecular dynamics simulations with detailed atomic models have been used to analyze the initial stages of protein unfolding. Results concerning possible folding intermediates and the role of water in the unfolding process have been obtained. The two types of theoretical approaches are providing information essential for an understanding of the mechanism of protein folding and are useful for the design of experiments to study the mechanism in different proteins.

Current Opinion in Structural Biology 1995, 5:58–73

Introduction

A very large number of distinct conformations exist for the polypeptide chain of which a protein molecule is composed. The protein spends most of its time in the native conformation, which spans only an infinitesimal fraction of the entire configuration space. Thus, the amino acid sequence of a protein must satisfy two requirements, one thermodynamic and the other kinetic. The thermodynamic requirement is that the sequence must have a unique folded conformation, which is stable under physiological conditions. The kinetic requirement is that the denatured polypeptide chain can fold into this conformation with reasonable speed. The dynamics of protein folding is not understood, although there have been many studies devoted to this subject [1–3]. A basic question concerns the mechanism by which a polypeptide chain is able to fold rapidly (on a millisecond to second timescale) to the stable native state, despite the very large number of conformations that exist for the chain (Levinthal paradox [4]). Recently, it was demonstrated that the protein folding problem is NP-hard [5–8]. It was shown [5,6] that an essential element of the computational complexity of protein folding is the presence of long-range interactions, which lead to the cooperative character of the folding transition. This result made possible a reformulation of the Levinthal paradox [6] that clarifies certain problems, with the original statement [4] and with some claims for the resolution of the Levinthal paradox [9].

During the past few years, there has been a great increase in the level of interest in protein folding. This is due in part to the challenge of the human genome project and in part to the development of experimental methods that provide more details about the folding process. Structural information concerning the folding and unfolding reaction comes from stopped-flow kinetics and NMR experiments [10–14], circular dichroism [10,15], fluorescence life-time measurements [15], mass spectrometry [16], and mutation studies [17,18]. Most of the kinetic data have been limited to time periods of milliseconds or longer, with earlier events in the 'dead-time' of the experiment. One conclusion from the measurements is that an essential part of the folding process, the search for an ordered globule with many attributes of the native structure, is completed within the deadtime [13]. Thus, new methods for rapidly triggering the folding reactions are crucial for obtaining results concerning the critical events that take place during the first millisecond ([19]; M Gruberle, personal communication).

There is renewed optimism that the protein folding problem can be 'solved'. Many people, including physicists and mathematicians as well as chemists and biologists, are making contributions to the theory of protein folding (recently reviewed in [1–3,20–31]). This has led to the introduction of a wide range of theoretical approaches. The present limitations of computing power require that simplified models be used for simulations of the entire folding process, which starts with a random coil configuration and ends in the native state. This is to be contrasted with the studies of the native state itself, which has relatively small (2 Å) and rapid (subnanosecond) fluctuations. These small and rapid fluctuations have been examined in great detail with molecular

Abbreviations

BPTI—bovine pancreatic trypsin inhibitor; **HP**—hydrophobic/polar; **i**—amino acid residue; **MECS**—minimum energy, maximally compact structures; **T**—temperature.

 ISSN 0959-440X

dynamics simulations that employ potentials with a full atomic representation of the protein chain and the solvent [32].

Much less is known about the features of the potential surface governing the non-native portion of configuration space involved in protein folding. This includes a wide range of structures that may differ by tens of angstroms and be separated by significant energy barriers [3,33–35]. Thus, the times corresponding to the motions in the full conformation space are typically within the nanosecond to second range. The resulting separation of time and distance scales between fast local motions and slow larger-scale motions makes it possible to introduce two simplifying concepts that serve as the basis of much of the theoretical work on protein folding. The first is a temperature-dependent effective potential, or potential of mean force, and the second is a discretized description of the polypeptide chain. Both of these concepts are based on the idea of 'pre-averaging' the small-scale motions to obtain a 'coarse-grained' model that can treat a molecule on the time and distance scales on which protein folding occurs. This is the rationale for protein models that include only a subset of atoms [36] or describe the protein by its secondary structure segments the dynamics of which are simulated in continuous space [37–39]. The additional simplification of discretizing the conformational space is achieved by lattice models [98,139,142] that employ Monte Carlo algorithms to simulate the kinetics. Because of the coarse graining inherent in such models, the substates, which have been shown to play an important role in native state dynamics [20,32], are generally encompassed in a single lattice conformer.

Because of the paucity of detailed experimental information and the impossibility of complete atom-based simulations of protein folding, speculations in this area are widespread and often useful. Many models for folding have been proposed. They are based on theoretical considerations [33,35], on phenomenological constructs [31,40–42] and on computer simulations (e.g., [43–46]; see also several recent reviews [3,22,31]). The present review is concerned with theories and simulations of the process of protein folding and unfolding; it does not deal directly with methods for predicting the native structure (as presented in [28,37–39,47–54], for example) nor with the thermodynamics of protein stability [55–57]. We indicate what mechanisms are being considered, how they differ and how they are related, and what evidence supports one or another mechanism. Both simplified and all-atom models are described as appropriate. We emphasize recent computer simulations, as theoretical studies were reviewed recently in [3] and phenomenological models in [31].

Molecular dynamics simulations of protein folding

Molecular dynamics simulations with an atomic model and explicit representation of the solvent can be used for studying unfolding from the native state. To reduce the time scale of this reaction, which generally has a significant activation barrier [16,58], to that accessible by simulations, the studies have been performed at high temperatures [59–63], at extreme pH (A Caflisch, M Karplus, unpublished data), or with artificial energy terms that force the unfolding (PH Hünenberg, AE Mark, WF Van Gunsteren, abstract, Summer School on 'Understanding Protein Motions', Huddinge, August 1994; C Brooks III, personal communication).

Most simulations have used high temperatures to speed up the unfolding reaction. Temperatures in the range of 400 to 600 K have been employed. If the activation barrier for unfolding is 20 kcal mol^{-1} and the experimental denaturation temperature is 325 K, an Arrhenius model calculation leads to an acceleration of the unfolding reaction by about six orders of magnitude at 600 K relative to 325 K. Although the Arrhenius equation is not expected to be valid for such a complex reaction (as there may be, for instance, important frictional effects, and the activation energy may be temperature dependent), this suggests that the time for unfolding is reduced from the experimentally observed millisecond range to nanoseconds, a time scale attainable in present day simulations.

Reviews of much of the published and some unpublished all-atom simulations of the protein unfolding process have appeared recently in this journal [21] and elsewhere [22]. They include simulations of unfolding of α-lactalbumin [59], lysozyme [60], bovine pancreatic trypsin inhibitor (BPTI [61,62]), barnase ([22,63]; A Caflisch, M Karplus, unpublished data), apomyoglobin [64,65] and β-lactamase [66]. We do not review individual papers that were included in [21] but focus on certain points that can profit from discussion. Simulations are of interest primarily because they offer the possibility of providing much more detailed information than is available from experiment [32]. Because the simulations are approximate, however, it is essential to validate them by experiment. In contrast to comparisons of the native structure and its fluctuations [32], such comparisons are rather difficult for the unfolding reaction; for instance, it is not clear how the series of events happening in a nanosecond simulation are to be related to the structural data for transition states and intermediates observed in experiments on protein folding and unfolding. Thus, although we agree with Daggett and Levitt [21] that the time course of the structural changes observed in unfolding simulations is of great interest, it appears to us that the experimental evidence for their accuracy is still rather limited.

The role of solvent in protein folding and unfolding is one of the essential questions that can be examined by simulations. Such studies are of particular interest because of the lack of experimental information. Furthermore, as the solvent (consisting of the order of 10 000 atoms) consumes most of the computer time in any unfolding simulation (in barnase, for example, there are only about 1000 protein atoms), it behoves the simulator to examine what the water is doing. In a series

of simulations of the early stages of barnase unfolding at high temperature [22,63], it was shown that solvent plays a key role in the denaturation process. For the main α-helix in barnase, the role of the solvent is similar to that observed in earlier unfolding simulations for isolated helices [22,67–69]. The simulations show that an important element of the helix-unfolding transition is the replacement of an α-helical (i to i+4, where i is an amino acid residue) hydrogen bond by water hydrogen bonds through an intermediate involving a 3_{10} (i to i+3), or reverse turn, hydrogen bond. The exact coupling between helix unfolding and water insertion and its dependence on temperature are not yet clear, however.

Denaturation of a β-sheet was described for the first time in the barnase simulations [22,63]. Barnase has a central five-stranded β-sheet; the three inner strands were found to be more stable than the two outer strands. Solvation of the outer strands was observed to start by the distortion of the β-sheet hydrogen bonds, followed by the insertion of hydrogen-bonding water molecules between the strands. From analysis of a series of mutation experiments [17], the central stabilizing element of barnase appears to be a hydrophobic core between the β-sheet and the main α-helix. The denaturation of this hydrophobic core was studied in the simulations. It also involved significant solvent participation. Water molecules initially entered by binding to polar side chains. An increase in the accessible surface area due to the movement of the α-helix away from the β-sheet and water penetration were nearly simultaneous. Ultimately, the hydrophobic side chains in the center of the core became solvated. Solvating water molecules, some of which form cage structures around hydrophobic groups, tended also to be involved in hydrogen bonds to waters outside the core. An analysis of the electrostatic and van der Waals energy showed that more favorable protein–water interactions replaced the internal protein interactions during the denaturation process.

Because the temperatures used in such simulations are unrealistic, there are concerns as to whether the observed behavior corresponds to the actual denaturation process. For example, the barnase simulation was done at high temperature with the room temperature water density. For water at such a high pressure and high temperature, the diffusion constant is only slightly increased relative to that at room temperature (A Caflisch, M Karplus, unpublished data). However, it is possible that the high pressure forces water into available cavities. To test the importance of this effect, a simulation of the unfolding of barnase at low pH was performed (A Caflisch, M Karplus, unpublished data). It is known that low pH denatures barnase [70]; information about the low pH denatured structure has been obtained from NMR studies [71]. In a 450 ps simulation at 360 K, with all glutamate and aspartate residues and the C-terminal carboxyl group neutralized and with histidine doubly protonated, results very similar to those from the high-temperature simulation were obtained for the role of water in the denaturation process. Moreover, the overall structural changes in the two sets of simulations were very similar, though the unfolding process was significantly slower at the lower temperature. These results are in accord with the results of Daggett and Levitt who have varied the temperature in simulations of the unfolding of several proteins (unpublished data reported in [21]).

One of the few other simulations where the role of solvent is discussed, even if briefly, is a high-temperature study of the unfolding of BPTI ([61]; M Levitt, personal communication). Water entered exposed regions as the radius of gyration increased. This simulation was performed using the low water density corresponding to that of the high-temperature system. When the temperature and water density were set equal to the room-temperature values, rapid water penetration into the expanded BPTI structure was observed. In a very recent high-temperature simulation of the unfolding of a stable truncated form of chymotrypsin inhibitor 2 [72], the same methodology as in [61] was used. The solvation process appeared to be similar to that described for barnase [22,63], but details were not given. The transition state for unfolding was presumed to be populated after 230 to 250 ps of simulation time because more rapid structural changes occurred following that period and because the best agreement with mutation data on the transition-state structural characteristics [73,74] was obtained from that period; the difficulty of determining the transition state in complex systems like proteins was emphasized [72]. As found in experiments [17] and simulations [22,63] for barnase, disruption of the principal hydrophobic core and portions of secondary structure was also suggested as the essential step in unfolding of chymotrypsin inhibitor 2 [72].

It is not known from experiment when water penetrates the protein interior in the denaturation process, because direct measurements are difficult. It has been inferred that water enters rather late [75], though it has also been suggested that it is present in the molten globule [76]. Possible approaches to determining the interactions of water with protein residues during denaturation include photochemically induced nuclear polarization [77], nuclear Overhauser effect of water interactions with specific residues [78], chemical markers [79], and NMR of [^{19}F]-substituted amino acids [80].

In evaluating unfolding simulations, it is important to have some checks that the protein is stable under physiological conditions with the potential functions that are used in the simulation [21]. For barnase, for example, such stability was demonstrated by a 300 ps simulation at room temperature [22,63]. Of importance also is the fact that four buried water molecules, which had been omitted from the original structure, reached their correct position and that no other waters penetrated to the interior of the molecule.

Lattice models of proteins

Because of the lack of knowledge concerning the details of the mechanism by which a polypeptide chain

solves the Levinthal paradox, many different proposals have been made. Some of these have been examined by computer simulations. The most useful simulations employ highly simplified models of the lattice-bead type described briefly in the introduction. Their simplicity makes it possible to explore the accessible configuration space of the 'polypeptide' chain within a reasonable amount of computer time. It is necessary to ask, however, what such simple models can tell us about protein folding, as none of them represent proteins with sufficient accuracy to prove that a given folding mechanism is correct. Instead, they suggest possible mechanisms that can serve to resolve the Levinthal paradox in systems where long-range interactions are important [6]. The hope is that the conceptual results will stimulate experimentalists to think about the problem in new ways [81] and will aid in the design of experiments to test the proposed mechanisms [73,74,82].

In the lattice models, the protein chain is represented as a string of beads on a two-dimensional square lattice [83,84] or a three-dimensional cubic lattice [85]. The interactions between monomers, which provide the energy function for the Monte Carlo simulations, have been obtained from the random energy model [86,87], from the binary hydrophobic/polar model [84], or from statistical analysis of residue interactions in proteins [45,88]. In most of these models, the purpose is not to examine the folding of a particular amino acid sequence. Instead, the idea is to include the most essential features of proteins and to use the simulations to explore the general characteristics of possible folding mechanisms. Two essential features of proteins are the heterogeneous nature of the interactions, which arises from the presence of different amino acids, and the long-range nature of the interactions, which is due to the possibility that amino acids are close in space even though they are distant along the polymer chain. These two features, which can be included in lattice heteropolymer models, can lead to a unique ground state and a cooperative folding transition, two fundamental properties of real proteins that are required for their biological function [3].

As the model does not include side chains, the 'native' state is a compact globule with the native fold. Such structured globules may correspond to the experimentally observed molten globules, which are somewhat expanded relative to the native state, but preserve much of the backbone structure, and have side chains that undergo dihedral angle transitions [89–91]. Molten globules appear to be a late stage in the folding of some proteins and their formation involves resolution of the Levinthal paradox [89–91]. Thus, explorations with this type of model can give useful information concerning the folding mechanism. The final stages of the folding process, in which the system goes from a low-resolution structure to the true native state with tightly packed side chains, can be studied, in principle, by all-atom models of the type described in the previous section [49,92–94]. The suitability of lattice models for describing real proteins and the use of the Monte Carlo algorithms for studying the folding dynamics are discussed in [87,95,96].

To focus the discussion of lattice models, we first describe some recent results obtained for a 27-mer heteropolymer chain on an infinite cubic lattice [44,81,87]. We then compare the 27-mer results with those from other models based on statistical mechanics and on lattice simulations. This comparison does not show that one or another model is correct or incorrect; rather, it makes clear that there are significant differences in the mechanisms that have been proposed for protein folding, although there is some commonality in most of them. More experimental data are needed to determine which of the proposals, if any, are valid for proteins.

There are two distinctive aspects of the 27-mer study [44,87]. The first is that the lowest energy state (the native state of the "toy protein") for each sequence was known, so that one could determine whether a given Monte Carlo folding simulation actually found the native state. Most other simulations, with the exception of those based on shorter chains and restricted to two dimensions (for instance, [46,84]), did not determine the actual lowest energy state (see, for example, [97]). The second aspect of this study is that it is a real computer 'experiment'. Many people refer to computer simulations as 'experiments', even though they are not. They are based on a theoretical model, and the only difference from an analytic theory is that the equations involved are solved on a computer. In the 27-mer experiment, 200 random sequences were generated. A sequence is characterized by the interaction matrix for all pairs of beads with matrix elements chosen from a Gaussian distribution [34]. Each was subjected to ten folding simulations of up to 5×10^7 Monte Carlo steps. Thirty of the 200 sequences found the known native state in four or more of the 10 simulations (strongly folding sequences), whereas 146 sequences never found the native state (non-folding sequences); the others found the native state in one to three simulations. By examining the two limiting sets of sequences, strongly folding and non-folding, it was possible to determine the essential aspects that differentiated them. Somewhat to our surprise, no structural features distinguished the strongly folding sequences. Furthermore, the results did not support several models [40,41,43,86,98] that had been proposed for solving the Levinthal paradox in proteins. Instead, the only special attribute of the strongly folding sequences was the presence of a large energy gap between the native and the excited states; the non-folding sequences had no such energy gap. The presence of a large energy gap was shown to be a necessary and sufficient condition for folding in the 27-mer model. It was shown to be necessary because no sequence without such a minimum folds to the native state, and to be sufficient because all sequences wih such a minimum do fold. In the detailed calculations, the energy gaps from the enumerable states of a $3\times3\times3$ cube were used; comparison with an analysis of the entire conformational space based on Monte Carlo sampling [87] gave corresponding results. Further,

there were 11 out of the 200 sequences that have their minimum outside the fully compact set. None of these satisfied the energy condition nor did they fold repeatedly to either the lower fully compact state or the lowest energy state found from Monte Carlo simulations. Thus, these sequences confirmed and generalized the folding criterion based on the fully compact states.

The importance of an energy gap gains some support from off-lattice C_α and C_β bead models with statistical effective potentials used to study the inverse folding problem [99–101]. It has been found that the correct structure for a given sequence is significantly lower in energy than the distribution of energies obtained from threading the sequence through a 'superprotein' constructed from known protein structures. Furthermore, the superprotein-derived structures tend to have a Gaussian distribution of energies, in accord with the energy spectrum of the 27-mer model and random-energy heteropolymers [33,34]. Energy level diagrams for lattice proteins, which compare the characteristics of systems with and without energy gaps and were first described in [85] and subsequently employed in [44,87], are now being widely used for discussions of protein stability and folding [29,102,103]. An interesting application of the energy gap concept focuses on small fragments and correlates the energy gap between the lowest and higher energy conformations of these fragments with their role as early folding units [153].

A reason for the correlation between folding and stability is that significant portions of the potential energy surface of the model system are 'rugged'. In particular, the random collapsed state that is sampled in the three-stage random search (3SRS) mechanism (see below) is a multiminimum surface on which the search for the native state requires surmounting many intervening barriers. This can be done on a reasonable time scale only if the folding temperature is sufficiently high for there to be a significant probability of overcoming such barriers. However, at a high temperature, the majority of the random sequences have a ground state that is not stable; i.e. the Boltzmann probability of being in an excited state is too large unless a sizeable energy gap separates the native state from the excited states. As the temperature at which the folding simulations are done is near the midpoint of the thermodynamic transition temperature between the native and denatured states, the simulation temperature is high enough to overcome the barriers only for the strongly folding sequences with a particularly stable ground state. The transition temperature for the non-folding sequences is so low that the 27-mer gets trapped in a metastable well. This qualitative argument explains only why the pronounced energy minimum is necessary for folding. The explanation for why it is also sufficient is provided by the 3SRS mechanism discussed below. It is likely that the necessity of the energy gap condition is general for reasonable lattice models and for real proteins; its sufficiency may be of more limited applicability.

Each simulation for a given non-folding sequence ended in a different state. This negates the suggestion that the native state is likely to be a metastable state, as it would require that a given sequence always folded to the same metastable state in this model. Honeycutt and Thirumalai [104], amongst others, have suggested that the native states of proteins are metastable states, although their off-lattice results [104], like those from the 27-mer lattice simulations [87], do not support this hypothesis. Experiments indicate that α-lytic protease and serpins can exist in different but similar conformations, which are of comparable stability under the same conditions [105–107]; in the case of α-lytic protease, a molten globule like state appears to be the trapped configuration. This opens the possibility of kinetic control of the folding reaction for some proteins. Also of interest is the suggestion that prions may function by having more than one state with similar energies [108].

The importance of temperature in the protein folding reaction and the relation between an energy gap and folding, which are clearly demonstrated in the 27-mer simulations [44,87], have been discussed previously. Based on insightful statistical mechanical arguments and spin glass theory, Bryngelson and Wolynes [33,35] suggested that two temperatures need to be considered in determining the folding properties of a sequence. One is the folding temperature, which corresponds to the midpoint of the thermodynamic transition between the native and denatured states, and the other is the glass transition temperature, which corresponds to the temperature below which the chain is frozen into a random low-energy conformation because it does not have enough energy to overcome the barriers separating such conformations. Thus, the temperature at which the sequence folds must be higher than the glass transition temperature. Furthermore, they showed that random sequences do not satisfy this condition and so would be likely to be trapped in metastable states [33,35]. Bryngelson and Wolynes introduced specific biases toward the native state to make folding possible. The existence of such biases on the entire potential surface corresponds to the principle of "minimum frustration" [33,35], which is closely related to the "consistency" or "harmony" principle proposed by Go and Abe [98]. One way of introducing the necessary bias is by the use of associative-memory Hamiltonians [109–113], which have been employed successfully in a variety of applications; e.g. Goldstein *et al.* [109] showed that the ratio between the folding and glass transition temperatures, the maximization of which was assumed to lead to faster folding, is proportional to the ratio of the energetic separation of the native state from the denatured states and the range of energies corresponding to the denatured states. Bryngelson and Wolynes also suggested [114] that there is a large degree of overlap between the structures of the transition state and the native state. It is clear from this paragraph that the theoretical studies of Wolynes and coworkers presaged a number of the results of the 27-mer simulations [44,87]. What is important about the 27-mer study is that it provided the first demonstration that the energy gap condition and a detailed mechanism for folding could be found *a posteriori* in a computer experiment, rather than having to be

introduced *a priori* into a model to achieve active folding, as in [33,35,98,109,114].

The 27-mer simulations with interactions based on the random energy model [44,87] demonstrated that sequences which have a large energy gap (and thus with a thermodynamically stable native state) also can find the native state (and as such, are kinetically viable). Of course, this does not imply that such a relationship is present for all potential surfaces. As pointed out previously [32,33,87], surfaces of the 'golf course' type (flat, with a deep, highly localized global minimum) do have a large energy gap and so would be thermodynamically stable, but not kinetically viable. Conversely, potential energy surfaces with a smooth approach to a minimum (funnels) would not have a glass transition temperature and so could fold even without a large energy gap; they would be kinetically viable but not thermodynamically stable. We believe that neither the golf course nor funnel-type surfaces are likely to correspond to real proteins and that the model illustrated by the 27-mer simulations is more probable, as it leads to a simple evolutionary selection criterion for stability and folding (see below).

The insights concerning the role of the temperature and the energy gap provided by the 27-mer computer experiment and theoretical analyses are of considerable interest. Nevertheless, such correlations do not, in themselves, provide a specific mechanism by which a system resolves the Levinthal paradox. For this purpose, further examination of the results for the strongly folding sequences in the 27-mer is required. Such an examination is possible as all the information, at any level of detail, is preserved in the Monte Carlo trajectories. This is a particularly useful aspect of computer experiments in contrast to laboratory experiments. Analyses of individual trajectories for both strongly folding and non-folding sequences, as well as the calculated density of states and reaction profiles [44,87] demonstrated that a three step random search (3SRS) mechanism was involved. In this mechanism, folding starts by a rapid collapse (approximately 10^4 Monte Carlo moves) from a random-coil state to a random semicompact globule. This involves a simple 'downhill' process, with the energy decreasing monotonically as the number of contacts increases. As most of the contacts are non-native, the semicompact globule is 'misfolded'. The chain then proceeds by a slow, rate-determining search through the semicompact states to find a transition state from which it folds rapidly to the native state. The elements of the 3SRS mechanism that lead to the resolution of the Levinthal paradox are the reduced number of conformations that need to be searched in the semi-compact globule ($\sim 10^{10}$ versus $\sim 10^{16}$ in the random coil) and the existence of many (10^3) transition states. Thus, a random search through the semicompact states involving 10^7 ($=10^{10}/10^3$) Monte Carlo moves is sufficient to find one of the transition states. The search is random in the sense that, over many trajectories, the microscopic states are occupied according to their Boltzmann probabilities and that there are many microscopic states with comparable Boltzmann probabilities at the random-coil, random globule and transition-state stages of folding.

As there are many transition states, there are many ways of reaching the native state. However, the multiplicity of transition states does not imply that their structures are dissimilar. In fact, all of the transition states have more than 80% of the native contacts, although no specific set of contacts is present in all states. Fersht *et al.* [82] recently used a mutational analysis of the transition state for the folding of barnase and of a stable fragment of the chymotrypsin inhibitor 2 to attempt to determine whether the multiplicity of transition states in the 3SRS mechanism was consistent with the observed behavior of these proteins. According to Fersht *et al.* [82], Baldwin [81] proposed that the 3SRS mechanism [44] presented "an immediate challenge for experimentalists to determine whether or not the folding of real proteins has a unique transition state". From their results, Fersht *et al.* [82] concluded that "the very first stages of folding do have multiple pathways but that these funnel into a single pathway in the later stages for many proteins" [18]. This suggestion and the experimental results in [82] are consistent with the 3SRS mechanism. The essential point, already mentioned, is that the transition states are close to the native state and have 80% of the native contacts. Thus, at the resolution of the mutation experiments, the multiplicity of transition states in the 3SRS mechanism corresponds to a unique transition state.

In the analysis by Fersht and coworkers, emphasis is placed on the structure of the transition state as probed by the dependence of the folding and unfolding rates on mutations. Some mutations were found to have a larger effect on the rates of folding than on the stability of the native state. Such detailed effects (i.e. a relatively small variation in rates) may not be decribed by the 27-mer model in its present implementation.

In the 27-mer model, the non-folding sequences have semicompact globules and transition states that are similar in energies and numbers to those of the folding sequences. The only significant difference is that the native states of the non-folding sequences are less stable than those of folding sequences. Thus, folding of the non-folding sequences would have to occur below the glass transition temperature, where the chains are frozen in random metastable conformations and never reach the native state, or reach it very slowly.

The kinetic behavior of the strongly folding sequences of the 27-mer was found to be a single exponential process over a wide temperature range above the glass transition temperature; i.e. the kinetics obey a unimolecular rate equation with a rate coefficient k independent of time (A Šali, E Shakhnovich, M Karplus, unpublished data; see Note added in proof). This is true even though the rate-determining step involves a search of the rugged surface corresponding to the collapsed state. The results justify the use of a relatively high temperature in the folding experiments so as to convincingly determine non-folding sequences and save computer time. Examination of the

variation of ln k versus 1/T, where T is the temperature, shows that the temperature dependence of the rate constant is complex: at low temperatures the rate increases as a function of temperature, whereas at high temperatures the rate decreases as a function of temperature. This behavior corresponds to a reaction that is energy dominated at low temperatures (i.e. crossing energy barriers is rate limiting) and entropy dominated at high temperatures (i.e. the configuration space that is accessible increases with temperature and requires a longer search to reach the transition state). A corresponding temperature dependence of the folding rate, which in some models is biphasic [115], has been found in a number of lattice simulations [84,95,115–117]. Experimental studies of protein folding [14,75] also show such a non-Arrhenius temperature dependence of the rate constant. However, an interpretation of the experimental result was based on solvation effects [75], although there is no direct evidence for this.

The first stage of the 3SRS mechanism, collapse of an expanded random coil into a random semicompact state, has been proposed as an early step in protein folding by many people. In a thermodynamic model for the stability of the native state [27,55], Dill [55] recognized that the greatly reduced number of conformations in the semicompact state would contribute to the solution of the Levinthal paradox, but he also pointed out that it was not sufficient. Such a collapse was found to be an early event in several other lattice model simulations of protein folding [46,86,95,116,118,119]. Based on experimental studies of protein folding, an early collapse to a random state has been suggested [13,15]; collapse to a state with a hydrophobic core has been observed [154]. An important question concerns the relative order of secondary structure formation and collapse [13]. This has not been resolved experimentally, primarily because the time scale of these events has not been accessible (see above and [155]). Recent studies on protein folding mediated by chaperone proteins suggest that compact disorganized globules with excess hydrophobic surface, perhaps like the compact random globules of the 3SRS mechanism, may be the species that are prevented from aggregating and are protected from proteolysis [120–122]. The 3SRS mechanism also suggests that chaperonins may provide a container that could restrict the volume of the random globule and thus reduce the number of conformations that need to be searched to find a transition state.

The second stage of the 3SRS mechanism, a random search through semicompact globule configurations to find a transition state similar to the native state, is related to the slow rearrangements found for homopolymer [117] and hydrophobic/polar heteropolymer chains [95]. In those two studies, the chain dynamics on the square lattice was modeled by the transition matrix approach. A detailed description of the energy barriers and energy landscapes was given. The difference from the 3SRS mechanism is that, in [95], the transition states are described as expanded high-energy states lying between the compact non-native conformations and the native state, whereas in the 3SRS mechanism, the transition state comprises most of the states with at least 80% of the native contacts. Furthermore, although some of the microscopic transition states in the 3SRS mechanism may be local maxima in energy, the transition state region as a whole is not a maximum in the energy, but a maximum in the free energy [44,87]. In fact, the search for the transition states in the 3SRS mechanism appears to be rate limiting for entropic, and not energetic, reasons, at least at higher temperatures.

The 3SRS mechanism is different from the funnel model proposed by Onuchic and co-workers [43]. The funnel hypothesis assumes that sequences fold because they have a single large folding funnel leading to the native conformation, and that non-folding sequences do not fold because they have multiple pathways leading to several conformations. Evidence for a folding funnel was obtained by using a random 27-mer model on a cubic lattice and solving the transition matrix equation for the conversions between the maximally compact states of one folding sequence and one non-folding sequence [43]. The funnel for the folding sequence arose from the use of sequences whose interaction matrix is highly correlated with that of the native state (PE Leopold, personal communication), as in models of the type proposed by Go and Abe [98]. This contrasts with the 3SRS mechanism, where there is no special correlation for the folding sequences [87]. Another difference is that maximally compact states rarely occur on the folding pathway in the random 27-mer model [44,87]; the native state is one of the first maximally compact states reached in the Monte Carlo simulations. Moreover, above the glass transition temperature, the folding sequences in the random 27-mer tend to find their native states only marginally faster (not by more than an order of magnitude as in the funnel hypothesis) than the non-folding sequences [44,87,116].

Apparently discarding the original funnel hypothesis, Socci and Onuchic [116] have recently made a study of the random 27-mer model that is very similar to that in the original 3SRS model analysis [44,87], except for the use of hydrophobic/polar chains rather than random interactions. For six sequences, results similar to those in [87] were obtained, indicating that the choice of the interaction matrix does not strongly influence the general folding behavior. The folding rate was determined as a function of temperature and it was found that folding occurs only in a limited range above the glass transition temperature, which is again similar to the 3SRS results. A two-state process involving a fast collapse into a random globule followed by a slower rearrangement into the native state was described for the sequences that had a stable native state. Sequences with an unstable native state, referred to as "frustrated" by Bryngelson and Wolynes [33,35], did not fold. The character and number of the transition states, and the nature of the search process, all of which play an important role in the 3SRS mechanism, were not described.

In addition to the early suggestion that the native states of proteins are metastable (see above) [104], and off-lattice

folding simulations of a designed β-barrel protein [102], Camacho and Thirumalai [46] have made Monte Carlo folding simulations with the hydrophobic/polar model and 15-mer sequences on a square lattice. Two thermodynamic transition temperatures were introduced: a temperature below which there is a rapid collapse to a more compact random globule, and a lower temperature for which rearrangements lead to the native state. It was argued that many pathways existed for finding a small number of transition states that had much of the native structure. A three-stage mechanism was mentioned in the abstract but not elaborated in the paper. The mechanism considered is very different from that found in the random 27-mer model. In contrast to the 3SRS mechanism, where the rate-determining step is the progression from the random semicompact globule to the transition state with significant native structure, the search for intermediates with significant native structure is fast in the analysis by Camacho and Thirumalai, and the crossing of an energy barrier in the transition state region, which is not clearly defined, is slow.

In a subsequent study [123], the minimum energy, maximally compact structures (MECS) were enumerated for the same lattice model. It was found that the number of MECS was small and did not increase with chain length. It was again suggested that folding proceeds in three stages: random coil→random globule→MECS→native state. The last, rate-limiting, step was assumed to be restricted to a sampling of MECS and to involve the search for a transition state that had significant native structure. No simulation data was given to support this conclusion. Moreover, the MECS model is unlikely to be correct because folding occurs only above the glass transition temperature. The number of thermodynamically relevant (semicompact) states in a 27-mer on a cubic lattice is about 10^{10} [44], which is much larger than the small subset of 103 346 maximally compact states corresponding to the MECS. Thus, the search will involve mainly the semicompact states. Also, as mentioned above, folding in the random 27-mer model proceeds directly to the native state without a search of other maximally compact states. Finally, the MECS states are likely to have energy barriers between them that are on the order of N kT, where N is the number of monomers, so interconversion between them is likely to be extremely slow. It is likely that MECS correspond to frozen states, not to viable folding intermediates [102,123].

In their studies of the mechanism of folding, Dill and coworkers [84,95] have focused on two-dimensional square lattices with hydrophobic/polar interactions. These are probably less realistic than three-dimensional random energy models for studying the kinetics of folding and tend to have highly degenerate ground states; the latter problem can be diminished by using only the subset of sequences with unique ground states. The advantage of the model is that for very short sequences, of up to 15 or so residues in two dimensions, it is possible to do an enumeration of the entire configuration space [124]. A consequence of this is that for such sequences [46,84,95] simulations use more Monte Carlo moves to find the minimum energy state than would be needed for a random search. This means there is no Levinthal paradox. For a 13-mer on a square lattice [46,84], for example, there are only $\sim 4 \times 10^4$ conformers, and 10^5 or more Monte Carlo steps were required to find the native state [84]. Consequently, the relation of such studies to the kinetics of protein folding is not clear.

The first paper on the folding mechanism by Dill and co-workers [84] described three regimes for the kinetics as a function of the magnitude of the attractive hydrophobic–hydrophobic contact energy for a single sequence; hydrophobic–polar and polar–polar contacts were deemed to have zero energy. They found that for weak attraction folding was slow because the molecule searched randomly through the large ensemble of open conformations, whereas for strong attraction the molecules became trapped in local minima with non-native conformations. For intermediate attraction, folding could be fast because the driving force was sufficient to direct the molecule toward a small ensemble of compact states and yet the minima were not deep enough to slow this process. As the important variable is the contact energy divided by the temperature, this result corresponds to the 3SRS mechanism, where there is freezing below a certain temperature, successful search above that temperature as long as the search is restricted to the compact states, and finally, essentially no folding at an even higher temperature because there is no collapse to sufficiently compact states for an effective search.

Chan and Dill recently used a transition matrix approach to Monte Carlo kinetics to describe in great detail the folding trajectories of homopolymers [117] and heteropolymers [95]. They suggested [95] that folding is rate-limited by energy barriers and has an approximate Arrhenius-type dependence on the energy of the highest barrier; this corresponds to the low-temperature folding regime in the random 27-mer model simulations (see above). Another result of interest is that the absolute folding time, in terms of number of moves, is quite sensitive to the move set. This may be a consequence of the poor behaviour of the Monte Carlo algorithm in this case; i.e., as pointed out above, more Monte Carlo steps are required than the total number of conformers to be searched. They also comment that "sequences with larger energy gaps should be both stable and fold quickly", though this statement is not a direct consequence of their simulation results.

The 3SRS mechanism is expected not to be directly applicable to protein chains with more than about 80 residues. The folding time of a 3SRS system increases approximately exponentially with chain length, as the number of semicompact states increases faster than the number of the transition states. It is likely, however, that proteins existing early in evolution were small enough to fold according to the 3SRS mechanism [125]. Because the pre-biotic and early biotic environments were hot, unusually thermostable proteins were required, such as those found in the most primitive bacteria that live at temperatures as high as 380 K [126]. In such environ-

ments, the stability condition would have required a native state that corresponded to a very deep energy minimum, for which the folding problem would have been solved simultaneously.

As evolution progressed, larger proteins evolved. One possibility for the folding of such longer chains is that the randomness of the second stage of the 3SRS mechanism is reduced by the presence of native-like secondary structure; there is experimental evidence that this occurs in some proteins [31]. Lattice simulations of 125-mers on a cubic lattice (A Dinner, A Šali, M Karplus, unpublished data) have demonstrated the feasibility of such a folding mechanism. Local contacts would then play an important role in guiding the folding. A related type of reduction of the space that has to be searched is embodied in the hydrophobic zipper hypothesis [42,127]. In this model, hydrophobic contacts are assumed to act as constraints that bring other contacts into spatial proximity, which would then further constrain and 'zip up' the next contacts, and so on. This mechanism could involve compact intermediates that have much secondary, but little native tertiary, structure. Folding paths were predicted [42,127] for several real protein sequences approximated by the hydrophobic/polar model; although the predicted states are neither global minima nor actual native states, they do have very low energies. The configuration space reduction arising from hydrophobic zippers and/or hydrogen-bonded secondary structure is likely to play a role in the folding of real proteins.

Another possibility for reducing the randomness of the search in the second stage of the 3SRS mechanism is provided by the existence of a stable nucleus. Such a nucleation mechanism has been observed in folding simulations for sequences of varying lengths designed by Monte Carlo optimization in sequence space to have a large energy gap between the native and the excited states [45,115]. A statistical mechanical analysis of this sequence selection procedure and its relation to folding has been given [119,128–130]. In the simulations, most sequences collapsed rapidly to a random globule, then spent most of the time in finding the nucleus, from which there was rapid folding to the native state. The last step was fast even when the temperature was decreased below the glass transition temperature, indicating that the potential surface was smooth in this part of the configuration space. For 36-mer designed sequences with 48 native contacts [45], nuclei consisting of between seven and nine strongly attractive short- and long-range native contacts were found to be necessary and sufficient for rapid folding to the native state. As a consequence, the transition state in the nucleation mechanism is much less similar to the native state than is the transition state in the 3SRS mechanism. The relation of this result to the observation that at least some proteins [17,18,82] have a transition state that resembles the native state requires further investigation. The role of a pronounced global energy minimum is, in part, similar to that in the 3SRS mechanism. It makes the native state sufficiently stable to allow folding to proceed at temperatures above the glass transition temperature. In addition, an extremely stable native state appears to increase the chance of obtaining a small spatially compact unit of especially attractive native contacts (i.e. a nucleus) through sequence space optimization. No such nucleus was found in the random folding 27-mers ([44,87]; A Šali, E Shakhnovich, M Karplus, unpublished data). Thus, existence of a nucleus may be a consequence of the sequence design procedure.

From a recent mutational analysis of a stable chymotrypsin inhibitor fragment (64 amino acids) and barnase (110 amino acids), Otzen *et al.* [73] have suggested that the latter has a more complex folding mechanism. Barnase folding appears to involve a rearrangement of the semicompact globule that leads to the formation of organized secondary structure prior to the appearance of the native state. This is conceptually in accord with the extrapolation of the 3SRS mechanism to larger proteins.

Several other studies have used lattice models to address problems related to the mechanism of folding. One study [124] used 16-mers on a square lattice and made a rigorous calculation of the phase diagram as a function of temperature and the average attraction. The phase diagram provides important information about the stable regions of phase space that are likely to be involved in the folding process. Other work includes the exploration of the geometric 'forces' of protein structural organization [131,132], the description of the interactions involved in denaturation and in denatured states [133,134], sampling of the protein conformations to determine thermodynamic functions [97,135,136], sampling of compact conformations to determine the lowest energy structures [47,51] and the exploration of the role of side-chain packing in protein stability and structure [137]. Also, the properties and possible origins of protein secondary structure were examined by Socci *et al.* [138] by use of a model with a single bead per residue in continuous space; they concluded that the volume restraint by itself does not force the formation of secondary structure, in disagreement with earlier lattice studies [139,140]. Additional results are presented in [141], where it is concluded that compactness does increase the stability of the secondary structure, although the absolute amount of secondary structure depends strongly on its definition.

Lattice models of real proteins

More elaborate lattice models use interactions between monomers or atoms that are meant to describe those existing in proteins [142]. A wide range of lattices have been employed [143]. Some of these are very complex, with up to 90 or so possible moves at each lattice point. They are sufficiently fine grained to give a high-resolution representation (1 Å) of protein X-ray structures. Although these lattices cannot be examined as exhaustively as the simpler models, they can be used both for elucidating some features of the folding process and for protein structure prediction. Their primary applications have been in the prediction field. Here, we summarize only the simulations that also consider the nature of the folding transition.

Refined lattice models and associated Monte Carlo schemes for protein structure prediction by folding simulations have been described in [48] and applied in [144]. The interaction potentials are based primarily on the analysis of residue contacts in known protein structures. Relative to earlier work [96], new terms for cooperativity in hydrogen-bond formation and side-chain packing were introduced. The method and its applications are reviewed in [145]. Exhaustive Monte Carlo lattice folding simulations of four-helix bundles were described [146] and the balance between secondary and tertiary interactions was explored. The protein folding mechanism was described as an on-site assembly of secondary structure elements; early intermediates with some native secondary structure appeared first, followed by formation of a molten globule and subsequent freezing of its secondary structure and side chains. The dependence of the details of the mechanism on the nature of the Monte Carlo moves deserves further investigation [31].

A backbone α-carbon model of a protein on a cubic lattice was used and Monte Carlo simulations were employed for folding of eight small proteins with statistical contact potentials [88] and a surface term [118,147]. The transition from a random chain to low-energy compact states with about 50% of the native contacts was examined. The first event was a collapse into a semicompact state without significant similarity to the final low-energy state. This was followed by chain repacking that led to progressively lower values of the energy. Chain repacking involved a highly cooperative interplay between the formation of local and non-local interactions. The transition was characterized by rapid relaxation of shorter chain segments to form local contacts and slower relaxations of longer chain segments to form non-local contacts.

The 210 (diamond) lattice and an entropy sampling Monte Carlo algorithm were used to simulate folding of a 38-residue hydrophobic/polar model that included both local and global interactions [148]. The energy was used as the reaction coordinate and a first-order transition between the denatured and native states was found. The first-order transition arises because the entropy increases more slowly than the energy when the protein initially unfolds from the native state. The results agree with those in [44].

Other simplified models

Several studies have used a series of increasingly more detailed models in attempts to fold to the native state [49,93,94]. This hierarchical approach typically starts with a bead model with contact potentials and Monte Carlo dynamics and ends with an all-atom model with a detailed force field and molecular dynamics. The main emphasis of these models is in structure prediction, rather than in simulating the actual folding pathway. The leucine zipper structure of the transcription factor GCN4 (a coiled-coil protein) was predicted by using a combination of a Monte Carlo lattice simulation and full-atom molecular dynamics models [94]. In the best structures, the backbone atoms have a root mean square deviation of 0.8 Å from the X-ray structure. In [49], a simplified model was used to simulate the collapse of the chain followed by a search for low-energy conformations that made use of a succession of increasingly detailed models and potentials; the method was then applied to the avian pancreatic polypeptide.

Restraints to increase and decrease the volume were applied in simulations of unfolding and refolding of BPTI, respectively [149]. Intermediates were generated that had native-like secondary structure but were twice the size of the native state. They could be folded to the native state by minimizing a function consisting of the ECEPP/3 potential and certain constraints on the known shape of the native protein. From an open structure with native-like local structure, rapid folding to the native state occurred.

Several models assume that secondary structure forms first [31] and simulate the dynamics of secondary structure elements connected by flexible loops. Chelvanayagam *et al.* [150,151] predicted protein folding pathways at a coarse-grained level by assuming that the secondary structure elements that bury the most solvent-accessible surface area interact first; they compared their results with those from experiments on protein folding kinetics. Similar ideas have been used previously with Monte Carlo simulations of folding to predict folding pathways consisting of nucleation, propagation and diffusion/collision for several proteins [152]. Models for protein folding were developed in which the pre-determined secondary structure elements were used; applications to myoglobin were described [37–39].

Conclusions

During the past few years, protein folding has become a very active field of research. The investigation of lattice models as generic representations of proteins has led to important insights into the folding mechanism. One important conclusion is that for certain lattice models of relatively short polypeptide chains, a pronounced global energy minimum, which assures thermodynamic stability, is a necessary and sufficient condition for rapid folding to the stable native state. A detailed mechanism for how this leads to a resolution of the Levinthal paradox has been determined from the simulations for one model, and extensions of the mechanism to long chains have been suggested. Although such simulations make use of simplified models, the results are of considerable interest because, in combination with statistical mechanical analyses, they are providing a stimulus for new experimental studies of the mechanism of protein folding. Molecular dynamics simulations of the initial stages of protein unfolding performed with all-atom models have supplied information on the structure and solvation of partially unfolded species that can be compared with experimental data. Most lattice models apply to the folding process up to the final stage (which involves

side-chain packing), whereas the molecular dynamics simulations yield a detailed picture of the initial stages of unfolding up to the molten globule state. Thus, the combination of all-atom molecular dynamics and simplified lattice Monte Carlo models provides a way of spanning the entire time and distance scales of the folding/unfolding process.

As computing power increases, more sophisticated models of proteins will be used to simulate the protein folding process. Moreover, better potentials of mean force will be used to simulate the folding of real protein sequences. These will make it possible to improve the predictions. The success of the three-stage random search mechanism in finding the pronounced global minimum on a model potential surface suggests that, at least for small proteins, the bottleneck in structure prediction is the derivation of a suitable potential function, rather than the design of folding algorithms.

The new insights concerning the mechanism of folding need to be tested by experiment to determine what real proteins do. Proteins could use a combination of mechanisms that might vary with the external conditions. Additional experimental data on the process of folding, especially on the submillisecond time scale, are needed to compare with the results of the simulations.

Note added in proof

The study referrred to in the text as A Šali, E Shakhnovich and M Karplus, unpublished data has now been published [156].

Acknowledgements

We are grateful to Aaron Dinner, Eugene Shakhnovich, Georgios Archontis, Amedeo Caflisch, Oren Becker, and Lloyd Dimitrius for discussions concerning the protein folding problem. A Šali is a Fellow of The Jane Coffin Childs Memorial Fund for Medical Research. This work was supported in part by grants from the National Science Foundation and the National Institutes of Health. This article was largely prepared whilst AS was at the Department of Chemistry, Harvard University.

References and recommended reading

1. Creighton TE (Ed): *Protein folding.* New York: WH Freeman and Co; 1992.

2. Merz K Jr, Le Grand S (Eds): *The protein folding problem and tertiary structure prediction.* Boston: Birkhäuser; 1994.

3. Karplus M, Shakhnovich E: **Protein folding: theoretical studies of thermodynamics and dynamics.** In *Protein folding.* Edited by Creighton TE. New York: WH Freeman and Co; 1992:127–196.

4. Levinthal C: **How to fold graciously.** In *Mossbauer spectroscopy in biological systems. Proceedings of a meeting held at Allerton House, Monticello, IL.* Edited by Debrunner P, Tsibris JCM, Münck E. Urbana: University of Illinois Press; 1969:22–24.

This brief article is the only mention by Levinthal of what is now called the Levinthal paradox.

5. Ngo JT, Marks J: **Computational complexity of a problem in molecular-structure prediction.** *Protein Eng* 1992, **5**:313–321.

6. Ngo JT, Marks J, Karplus M: **Computational complexity, protein structure prediction, and the Levinthal paradox.** In *The protein folding problem and tertiary structure prediction.* Edited by Merz K Jr, Le Grand S. Boston: Birkhäuser; 1994:433–506.

A thorough review and comparison of various demonstrations that the protein folding problem is NP-hard. Implications for protein structure prediction and protein folding are discussed.

7. Unger R, Moult J: **Finding the lowest free energy conformation of a protein is an NP-hard problem: proof and implications.** *Bull Math Biol* 1993, **55**:1183–1198.

A lattice model is used to demonstrate that protein folding is an NP-hard problem. Implications for protein structure prediction, folding, and evolution are discussed.

8. Fraenkel AS: **Complexity of protein folding.** *Bull Math Biol* 1993, **55**:1199–1210.

A proof that protein folding is computationally an NP-hard problem.

9. Zwanzig R, Szabo A, Bagchi B: **Levinthal's paradox.** *Proc Natl Acad Sci USA* 1992, **89**:20–22.

10. Jennings PA, Wright PE: **Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin.** *Science* 1993, **262**:892–896.

A paper that demonstrates that a molten globule species that had been studied previously is on the kinetic folding pathway of apomyoglobin.

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

12. Baldwin RW: **Pulsed H/D exchange studies of folding intermediates.** *Curr Opin Struct Biol* 1993, **3**:84–91.

13. Dobson CM, Evans PA, Radford SE: **Understanding how proteins fold: the lysozyme story so far.** *Trends Biochem Sci* 1994, **19**:31–37.

14. Dobson CM, Evans PA: **Protein folding kinetics from magnetization transfer nuclear magnetic resonance.** *Biochemistry* 1984, **23**:4267–4270.

15. Elöve G, Chaffotte AF, Roder H, Goldberg ME: **Early steps in cytochrome *c* folding probed by time-resolved circular dichroism and fluorescence spectroscopy.** *Biochemistry* 1992, **31**:6876–6883.

16. Miranker A, Robinson CV, Radford SE, Aplin RT, Dobson CM: **Detection of transient protein folding populations by mass spectrometry.** *Science* 1993, **262**:896–900.

A pioneering study that demonstrates the utility of mass spectrometry as a complement to NMR pulsed-labelling experiments.

17. Fersht AR: **Protein folding and stability: the pathway of folding of barnase.** *FEBS Lett* 1993, **325**:5–16.

Experimental results on the folding kinetics and structural rearrangements of barnase are reviewed.

18. Fersht AR: **Pathway and stability of protein folding.** *Biochem Soc Trans* 1994, **22**:267–273.

19. Jones CM, Henry ER, Hu Y, Chan C, Luck SD, Bhuyan A, Roder H, Hofrichter J, Eaton WA: **Fast events in protein initiated by nanosecond laser photolysis.** *Proc Natl Acad Sci USA* 1993, **90**:11860–11864.

A technique for optically triggering the folding reaction in reduced cytochrome *c*, by photo-oxidation of the heme iron–carbon monoxide complex, makes it possible to observe early stages in folding.

20. Frauenfelder H, Wolynes PG: **Biomolecules: where the physics of complexity and simplicity meet.** *Physics Today* 1994, **47**:58–64.

A description of protein folding from the perspective of spin glasses. The paper discusses the complexities of the protein energy landscape and how the principle of 'minimal frustration' can result in rapid folding.

21. Daggett V, Levitt M: **Protein folding↔ unfolding dynamics.** *Curr Opin Struct Biol* 1994, **4**:291–295.

Molecular dynamics studies of protein unfolding are reviewed.

22. Caflisch A, Karplus M: **Molecular dynamics studies of protein and peptide folding and unfolding.** In *The protein folding problem and tertiary structure prediction.* Edited by Merz K Jr, Le Grand S. Boston: Birkhäuser; 1994:193–230.

A review of molecular dynamics studies of protein unfolding, with the emphasis on barnase simulations.

23. Dill KA: **Folding proteins: finding a needle in a haystack.** *Curr Opin Struct Biol* 1993, **3**:99–103.
A review of the protein folding problem with emphasis on search techniques for finding the global energy minimum on a given potential surface. Examples that appear to violate the thermodynamic hypothesis are discussed.

24. Chan HS, Dill KA: **The protein folding problem.** *Physics Today* 1993, **46**:24–32.
A review of various aspects of the protein folding problem, including forces guiding protein folding, simplified lattice models, homopolymer and heteropolymer collapse theories, and search strategies for finding the native state.

25. Rose GD, Creamer TP: **Protein folding: predicting predicting.** *Proteins* 1994, **19**:1–3.
A short statement on the prospects for predicting the three-dimensional structure of proteins.

26. Rose GD, Wolfenden R: **Hydrogen bonding, hydrophobicity, packing and protein folding.** *Annu Rev Biophys Biomol Struct* 1993, **22**:381–415.
The role of the three major forces determining protein stability and kinetics is reviewed.

27. Dill KA, Stigter D: **Modeling protein stability as heteropolymer collapse.** *Adv Protein Chem* 1995, in press.
A progress report on a statistical mechanical model for computing the stability of protein. It also describes the origins of folding cooperativity as a function of temperature, denaturants, pH, and salts. A discussion of normal versus cold denaturation and of conformational entropy is given.

28. Levitt M: **Protein folding.** *Curr Opin Struct Biol* 1991, **1**:224–229.

29. Abagyan RA: **Towards protein folding by global energy optimization.** *FEBS Lett* 1993, **325**:17–22.
A review of approaches for protein structure prediction. Detailed models of Monte Carlo optimization methods and protein energetics are described.

30. Skolnick J, Kolinski A, Godzik A: **From independent modules to molten globules: observations on the nature of protein folding intermediates.** *Proc Natl Acad Sci USA* 1993, **90**:2099–2100.
A short letter summarizing the protein folding intermediates observed in simulations and experiments. The formation of secondary structure, the molten globule, and side chain fixation are discussed.

31. Karplus M, Weaver DL: **Protein folding dynamics: the diffusion-collision model and experimental data.** *Protein Sci* 1994, **3**:650–668.
The diffusion-collision model is described and its quantitative and qualitative predictions are compared with experimental data and alternative models.

32. Brooks CL III, Karplus M, Pettit BM: *Proteins: a theoretical perspective of dynamics, structure and thermodynamics.* New York: John Wiley and Sons; 1988.

33. Bryngelson JD, Wolynes PG: **Intermediates and barrier crossing in a random energy model (with applications to protein folding).** *J Phys Chem* 1989, **93**:6902–6915.

34. Shakhnovich EI, Gutin AM: **Formation of unique structure in polypeptide chains. Theoretical investigation with the aid of a replica approach.** *Biophys Chem* 1989, **34**:187–199.

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

36. Levitt M, Warshel A: **Computer simulation of protein folding.** *Nature* 1975, **253**:694–698.

37. Monge A, Friesner R, Honig B: **An algorithm to generate low-resolution protein tertiary structures from knowledge of secondary structure.** *Proc Natl Acad Sci USA* 1994, **91**:5027–5029.
See [38] annotation.

38. Gunn JR, Monge A, Friesner R, Marshall C: **Hierarchical algorithm for computer modeling of protein tertiary structure: folding of myoglobin to 6.2 Å resolution.** *J Phys Chem* 1994, **98**:702–711.
This reference and [37] describe a method for simulating the dynamics of proteins represented by secondary structure segments connected by loops. The method is applied to myoglobin.

39. Calloway DJ: **Solvent induced organization: a physical model of folding myoglobin.** *Proteins* 1994, **20**:124–128.

40. Kim P, Baldwin R: **Intermediates in the folding reactions of small proteins.** *Annu Rev Biochem* 1990, **59**:631–660.

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

42. Fiebig KM, Dill KA: **Protein core assembly process.** *J Chem Phys* 1993, **98**:3475–3487.
The authors propose a hydrophobic zipper mechanism of protein folding. The model assumes that contacts during folding are made such that the entropy loss is minimized. This process is shown to lead to the global energy minimum for many short hydrophobic/polar sequences on the square lattice but is not as successful for longer chains.

43. Leopold PE, Montal M, Onuchic JN: **Protein folding funnels: a kinetic approach to the sequence-structure relationship.** *Proc Natl Acad Sci USA* 1992, **89**:8721–8725.

44. Šali A, Shakhnovich EI, Karplus M: **How does a protein fold?** *Nature* 1994, **349**:248–251.
The folding mechanism for a random 27-mer heteropolymer sequence on the cubic lattice is described. A three-stage random search mechanism explains why the pronounced global energy minimum is necessary and sufficient for rapid folding. Evolutionary implications of the model are discussed.

45. Abkevitch VI, Gutin AM, Shakhnovich EI: **Specific nucleus as the transition state for protein folding: evidence from the lattice model.** *Biochemistry* 1994, **33**:10026–10036.
Folding of 36-mer sequences designed to have a pronounced global energy minimum is simulated by Monte Carlo dynamics on a cubic lattice. It is shown that the formation of a specific nucleus leads to subsequent rapid folding to the native state.

46. Camacho CJ, Thirumalai D: **Kinetics and thermodynamics of folding in model proteins.** *Proc Natl Acad Sci USA* 1993, **90**:6369–6372.
Random hydrophobic/polar sequences of 15 beads on the square lattice are subjected to Monte Carlo folding simulations. Two transition temperatures are found: the first corresponds to a collapse from a coil to a (random) compact phase, and the second (lower) temperature corresponds to the rearrangements leading to the native state.

47. Hinds DA, Levitt MA: **A lattice model for protein structure prediction at low resolution.** *Proc Natl Acad Sci USA* 1992, **89**:2536–2540.

48. Kolinski A, Skolnick J: **Monte Carlo simulations of protein folding. I. Lattice model and interaction scheme.** *Proteins* 1994, **18**:338–352.
This is a description of the latest version of the authors' lattice Monte Carlo folding algorithm based on a hierarchy of models of increasing detail. The potential of mean force, predominantly of statistical origin, contains several novel terms that facilitate the cooperative assembly of secondary structure elements and the cooperative packing of the side chains.

49. Liwo A, Pincus MR, Rackovsky S, Scheraga HA: **Prediction of protein conformation on the basis of a search for compact structures: test on avian pancreatic polypeptide.** *Protein Sci* 1993, **2**:1715–1731.
The idea that the search among the compact structures is easier than the search among all structures is used to develop a four-stage protein-folding simulation consisting of increasingly more detailed models and potentials. An approximately correct prediction of the native structure of a 36-residue polypeptide is obtained.

50. Aszódi A, Taylor WR: **Folding polypeptide alpha-carbon backbones by distance geometry methods.** *Biopolymers* 1994, **34**:489–505.
A new computational protocol was developed for calculating globular and compact protein structures consistent with hydrophobic restraints.

51. Hinds DA, Levitt M: **Exploring conformational space with a simple lattice model for protein structure.** *J Mol Biol* 1994, **243**:668–682.
A low-resolution lattice is used to enumerate all compact backbone conformations for small proteins. Using simple structural and energetic criteria, based on a database of known protein structures, lattice structures can be selected that have significant similarities to the known native structures.

52. Brower RC, Vasmatzis G, Silverman M, DeLisi C: **Exhaustive conformational search and simulated annealing for models of lattice peptides.** *Biopolymers* 1993, **33**:329–334.
Lattice models of short chains are enumerated and simulated annealing is used to estimate the time for finding the global minimum. The extrapolation of the optimization times to longer chains indicates that it is imperative to find better algorithms for minimizing longer chains.

53. Sun S: **Reduced representation model of protein structure prediction: statistical potential and genetic algorithms.** *Protein Sci* 1994, **2**:762–785.
A reduced model in continuous space is described and applied to folding melittin and avian pancreatic polypeptide.

54. Wallqvist A, Ullner M: **A simplified amino acid potential for use in structure predictions of proteins.** *Proteins* 1994, **18**:267–280.
Reduced model of a protein is described and applied to short proteins.

55. Dill KA: **Theory for folding and stability of globular proteins.** *Biochemistry* 1985, **24**:1501–1509.

56. Privalov PL, Makhatadze GI: **Energetics of protein structure.** *Adv Protein Chem* 1995, in press.

57. Lazaridis T, Archontis G, Karplus M: **The enthalpic contribution to protein stability: insights from atom-based calculations and statistical mechanics.** *Adv Protein Chem* 1995, in press.

58. Sosnick TR, Mayne L, Hiller R, Englander SW: **The barriers in protein folding.** *Nature Struct Biol* 1994, **1**:149–156.

59. Fan P, Kominos D, Kitchen DB, Levy RM: **Stabilization of alpha-helical secondary structure during high-temperature molecular-dynamics simulations of alpha-lactalbumin.** *Chemical Physics* 1991, **158**:295–301.

60. Mark AE, Van Gunsteren WF: **Simulation of the thermal denaturation of hen egg white lysozyme: trapping of the molten globule state.** *Biochemistry* 1992, **31**:7745–7748.

61. Daggett V, Levitt M: **A model of the molten globule state from molecular dynamics simulations.** *Proc Natl Acad Sci USA* 1992, **89**:5142–5146.

62. Daggett V, Levitt M: **Protein unfolding pathways explored through molecular dynamics simulations.** *J Mol Biol* 1993, **232**:600–619.
A detailed description of a high-temperature molecular dynamics simulation of bovine pancreatic trypsin inhibitor unfolding in solvent. The unfolding starts with expansion and an increase in the solvent accessibility, followed by turn dissolution and finally by secondary structure unfolding.

63. Caflisch A, Karplus M: **Molecular dynamics simulation of protein denaturation: solvation of the hydrophobic cores and secondary structure of barnase.** *Proc Natl Acad Sci USA* 1993, **91**:1746–1750.
Details high-temperature molecular dynamics simulations of barnase unfolding. The solvation of a β-sheet and a hydrophobic core during unfolding is described.

64. Brooks CL III: **Characterization of "native" apomyoglobin by molecular dynamics simulation.** *J Mol Biol* 1992, **227**:375–380.

65. Tirado-Rives J, Jorgensen WL: **Molecular dynamics simulations of the unfolding of apomyoglobin in water.** *Biochemistry* 1993, **32**:4175–4184.
A molecular dynamics simulation is applied to study unfolding of apomyoglobin. The structures obtained at 85°C display properties similar to those of molten globules. The results suggest explanations for the effects of several mutations at the helix interfaces and show a correlation between the stability of a region and the number of side chain–side chain contacts in that region.

66. Vijayakumar S, Vishveshwara S, Ravishankar G, Beveridge DL: **Differential stability of β-sheets and α-helices in β-lactamase: a high temperature molecular dynamics study of unfolding intermediates.** *Biophys J* 1993, **65**:2304–2312.
High temperature (200 ps, 600 K) and room temperature control (180 ps, 300 K) simulations of β-lactamase in water are compared. The simulation was done at constant volume, with the density of water at room temperature. The authors analyze denaturation of the secondary structural elements (helices and sheets), but do not discuss the role of solvent.

67. Daggett V, Levitt M: **Molecular dynamics simulations of helix denaturation.** *J Mol Biol* 1992, **223**:1121–1138.

68. Tirado-Rives J, Jorgensen WL: **Molecular dynamics simulations of the unfolding of an α-helical analogue of ribonuclease A S-peptide in water.** *Biochemistry* 1991, **30**:3864–3871.

69. Soman KO, Karimi A, Case DA: **Unfolding of an α-helix in water.** *Biopolymers* 1991, **31**:1351–1361.

70. Pace CN, Laurentz DV, Erickson RE: **Urea denaturation of barnase: pH dependence and characterization of the unfolded state.** *Biochemistry* 1992, **31**:2728–2734.

71. Arcus VL, Vuilleumier S, Freund SMW, Bycroft M, Fersht AR: **Toward solving the folding pathway of barnase: the complete backbone ^{13}C, ^{15}N, and ^{1}H NMR assignments of its pH-denatured state.** *Proc Natl Acad Sci USA* 1994, **91**:9412–9416.

72. Li A, Daggett V: **Characterization of the transition state of protein unfolding by use of molecular dynamics: chymotrypsin inhibitor 2.** *Proc Natl Acad Sci USA* 1994, **91**:10430–10434.
A high-temperature all-atom simulation of unfolding in low-density solvent. A possible approach to determining the structure of the transition state is given. The results obtained using this method agree very well with those produced by Fersht's group using a mutational approach (see [73]).

73. Otzen DE, Itzhaki LS, ElMasry NF, Jackson SE, Fersht AR: **Structure of the transition state for the folding/unfolding of the barley chymotrypsin inhibitor 2 and its implications for mechanisms of protein folding.** *Proc Natl Acad Sci USA* 1994, **91**:10422–10425.

74. Jackson SE, ElMasry N, Fersht AR: **Folding of chymotrypsin inhibitor 2. 2. Influence of proline isomerization on the folding kinetics and thermodynamic characterization of the transition state of folding.** *Biochemistry* 1993, **30**:10436–10443.

75. Sugihara M, Segawa S-I: **Characterization of the transition state of lysozyme unfolding I.** *Biopolymers* 1984, **23**:2473–2488.

76. Ptitsyn OB: **The molten globule state.** In *Protein folding.* Edited by Creighton TE. New York: WH Freeman and Co; 1992:243–300.

77. Kaptein R, Dijkstra K, Nicolay K: **Laser photo-CIDNP as a surface probe for proteins in solution.** *Nature* 1978, **274**:293–294.

78. Otting G, Wütrich K: **Studies of protein hydration in aqueous solution by direct NMR observation of individual protein-bound water molecules.** *J Am Chem Soc* 1989, **111**:1871–1875.

79. Ghelis C: **Transient conformational states in proteins followed by differential labeling.** *Biophys J* 1980, **32**:503–514.

80. Frieden C, Hoeltzli SD, Ropson IJ: **NMR and protein folding: equilibrium and stopped-flow studies.** *Protein Sci* 1993, **2**:2007–2014.

81. Baldwin RL: **Matching speed and stability.** *Nature* 1994, **369**:183–184.
A News and Views article exploring the relationship of the results of the lattice simulations in [44,87] to the thermodynamics and kinetics of the folding of proteins.

82. Fersht AR, Itzhaki LS, ElMasry NF, Matthews JM, Otzen DE: **Single versus parallel pathways of protein folding and fractional formation of structure in the transition state.** *Proc Natl Acad Sci USA* 1994, **91**:10426–10429.
This paper is the first experimental study that compares its results with those obtained from the 27-mer lattice model described in [44,87].

83. Lau KF, Dill KA: **A lattice statistical mechanics model of the conformational and sequence spaces of proteins.** *Macromolecules* 1989, **22**:3986–3997.

84. Miller R, Danko CA, Fasolka MJ, Balazs AC, Chan HS, Dill KA: **Folding kinetics of proteins and copolymers.** *J Chem Phys* 1992, **96**:768–780.

85. Shakhnovich EI, Gutin AM: **Implications of thermodynamics of protein folding for evolution of primary sequences.** *Nature* 1990, **346**:773–775.

86. Shakhnovich E, Farztdinov G, Gutin AM, Karplus M: **Protein folding bottlenecks: a lattice Monte Carlo simulation.** *Phys Rev Lett* 1991, **67**:1665–1668.

87. Šali A, Shakhnovich EI, Karplus M: **Kinetics of protein folding: a lattice model study of the requirements for folding to the native state.** *J Mol Biol* 1994, **235**:1614–1636.
This study addresses the question of the differences between the folding and non-folding sequences using a 27-mer random heteropolymer model on a cubic lattice and Monte Carlo dynamics. A pronounced global energy minimum is found to be a sufficient and necessary condition for a sequence to fold in this model.

88. Miyazawa S, Jernigan RL: **Estimation of effective interresidue contact energies from protein crystal structures: quasi-chemical approximation.** *Macromolecules* 1985, **18**:534–552.

89. Ptitsyn OB, Uversky VN: **The molten globule is a third thermodynamic state of protein molecules.** *FEBS Lett* 1994, **341**:15–18.
A paper that analyzes protein unfolding for a series of proteins and concludes that there can be two first-order transitions: one between the molten globule and the native state and the other between the molten globule and the unfolded state.

90. Ptitsyn OB: **Protein folding: hypotheses and experiments.** *J Protein Chem* 1987, **6**:273–293.

91. Peng Z-Y, Kim PS: **A protein dissection study of a molten globule.** *Biochemistry* 1994, **33**:2136–2141.

92. Skolnick J, Kolinski A, Brooks CL III, Godzik A, Rey A: **A method for predicting protein structure from sequence.** *Curr Biol* 1993, **3**:414–423.

93. Karasawa T, Tabuchi K, Fumoto M, Yasukawa T: **Development of simulation models for protein folding in a thermal annealing process. 1. A simulation of BPTI folding by the pearl necklace model.** *Comput Appl Biosci* 1993, **9**:243–251.
The first in a series of four increasingly detailed models for simulating the process of protein folding is described.

94. Vieth M, Kolinski A, Brooks CL III, Skolnick J: **Prediction of the folding pathways and structure of the GCN4 leucine zipper.** *J Mol Biol* 1994, **237**:361–367.
A hierarchical approach consisting of lattice Monte Carlo and molecular dynamics simulations is illustrated by predicting a coiled-coil structure. This is an example of a folding simulation that successfully combines coarse-grained lattice simulations and all-atom molecular dynamics. Cooperative creation of the hydrophobic core leading to side-chain fixation is observed.

95. Chan HS, Dill KA: **Transition states and folding dynamics of proteins and heteropolymers.** *J Chem Phys* 1994, **100**:9238–9257.
A transition matrix method developed for homopolymers is used to simulate the collapse of short hydrophobic/polar heteropolymers on a square lattice and to study their subsequent folding to the native state. A folding model is inferred that is similar to the three-stage random search model. The main difference is in the transition states, which are described as slightly expanded states of higher energy with no significant similarity to the native state.

96. Skolnick J, Kolinski A: **Dynamic Monte Carlo simulations of a new lattice model of globular protein folding, structure and dynamics.** *J Mol Biol* 1991, **221**:499–531.

97. O'Toole EM, Panagiotopoulos AZ: **Monte Carlo simulation of folding transitions of simple model proteins using a chain growth algorithm.** *J Chem Phys* 1992, **97**:8644–8652.

98. Go N, Abe H: **The consistency principle in protein structure and pathways of folding.** *Adv Biophysics* 1984, **18**:149–164.

99. Sippl M: **Calculation of conformational ensembles from potentials of mean force: an approach to the knowledge-based prediction of local structures in globular proteins.** *J Mol Biol* 1990, **213**:859–883.

100. Sippl M, Weitckus S: **Detection of native-like models for amino acid sequences of unknown three-dimensional structure in a data base of known protein conformations.** *Proteins* 1992, **13**:258–271.

101. Sippl MJ: **Boltzmann's principle, knowledge-based mean fields and protein folding. An approach to the computational determination of protein structures.** *J Comput Aided Mol Des* 1993, **7**:473–501.
A review of the author's approach to determining effective potentials from known protein structures. The applications discussed are validation of experimentally determined structures, data base searches for identification of sequence–structure pairs, sequence–structure alignments, and calculation of protein conformation from amino acid sequence.

102. Guo Z, Thirumalai D, Honeycutt JD: **Folding kinetics of proteins: a model study.** *J Chem Phys* 1992, **97**:525–535.

103. Totrov M, Abagyan R: **Detailed *ab initio* prediction of lysozyme-antibody complex with 1.6 Åaccuracy.** *Nature Struct Biol* 1994, **1**:259–263.
A Monte Carlo optimization of a detailed energy function in dihedral angle space is successfully applied to model a lysozyme–antibody complex.

104. Honeycutt JD, Thirumalai D: **The nature of folded states of globular proteins.** *Biopolymers* 1992, **32**:695–709.

105. Baker D, Sohl JL, Agard DA: **Protein-folding reaction under kinetic control.** *Nature* 1992, **356**:263.

106. Katagiri K, Okada K, Hattori H, Yano M: **Bovine endothelial cell plasminogen activator inhibitor. Purification and heat activation.** *Eur J Biochem* 1988, **176**:81-87.

107. Banzon JA, Kelly JW: **β-sheet rearrangements: serpins and beyond.** *Protein Eng* 1992, **5**:113–116.

108. Prusiner SB: **Chemistry and biology of prions.** *Biochemistry* 1992, **31**:12277–12888.

109. Goldstein RA, Luthey-Schulten ZA, Wolynes PG: **Optimal protein-folding codes from spin-glass theory.** *Proc Natl Acad Sci USA* 1992, **89**:4918–4922.

110. Friedrichs MS, Goldstein RA, Wolynes PG: **Generalized protein tertiary structure recognition using associative memory Hamiltonians.** *J Mol Biol* 1991, **222**:1013–1034.

111. Friedrichs MS, Wolynes PG: **Toward protein tertiary structure recognition by means of associative memory Hamiltonians.** *Science* 1989, **246**:371–373.

112. Sasai M, Wolynes PG: **Molecular theory of associative memory Hamiltonian models of protein folding.** *Phys Rev Lett* 1990, **65**:2740–2743.

113. Sasai M, Wolynes PG: **Unified theory of collapse, folding, and glass transitions in associative-memory Hamiltonian models of proteins.** *Phys Rev A* 1992, **46**:7979–7997.

114. Bryngelson J, Wolynes P: **A simple statistical field theory of heteropolymer collapse with application to protein folding.** *Biopolymers* 1990, **30**:171–188.

115. Abkevitch VI, Gutin AM, Shakhnovich EI: **Free energy landscape for protein folding kinetics: intermediates, traps, and multiple pathways in theory and lattice model simulations.** *J Chem Phys* 1994, **101**:6052.
The rate-limiting events in the creation of the nucleus are explored for designed 36-mer sequences on the cubic lattice. It is found that at high temperature the barrier is entropic and at that at low temperature it is energetic. This is caused by the search for the nucleus being among many relatively high energy states at high temperature, whereas it is among a smaller number of almost frozen states at low temperature.

116. Socci ND, Onuchic JN: **Folding kinetics of proteinlike heteropolymers.** *J Chem Phys* 1994, **101**:1519–1528.
Six 27-mer hydrophobic/polar chains on a cubic lattice were subjected to Monte Carlo simulations to explore their folding kinetics. The paper reproduces some elements of [87]. The authors found that the folding time depends on sequence and is related to the amount of energetic frustration in the native state. They also determined that there is a fast collapse into a random compact state that is followed by a slow rearrangement of the chain to the native state.

117. Chan HS, Dill KA: **Energy landscapes and the collapse dynamics of homopolymers.** *J Chem Phys* 1993, **99**:2116–2127.

A transition matrix method is used to simulate the collapse of short homopolymers on a square lattice. It is found that the kinetics depend strongly on the move set. Some pathways show two rates: a fast collapse to non-optimal compact states, followed by a much slower re-arrangement to maximally compact conformations.

118. Covell DG: **Lattice model simulations of polypeptide chain folding.** *J Mol Biol* 1994, **235**:1032–1043.
Folding of eight small proteins is studied on a cubic lattice with the simulated annealing Monte Carlo method, using Miyazawa and Jernigan's contact interactions and surface energy terms. Folding proceeds by rapid collapse followed by chain repacking, which involves a cooperative interplay between the formation of local and non-local interactions.

119. Shakhnovich EI: **Proteins with selected sequences fold into unique native conformation.** *Phys Rev Lett* 1994, **72**:3907–3910.
This application of the Monte Carlo sequence design method to 80-mer heteropolymers on a cubic lattice shows that a sufficient condition for folding of long designed sequences is that they have a pronounced global energy minimum.

120. Hartl FU: **Protein folding: secrets of a double-doughnut.** *Nature* 1994, **371**:557–559.

121. Braig K, Otwinowski Z, Boisvert DC, Joachimiak A, Horwich AL, Sigler AL: **The crystal structure of the bacterial chaperonin GroEL at 2.8 Å.** *Nature* 1994, **371**:578–586.

122. Fenton WA, Muslin AJ, Kikuchi A, Martin JA, MacNicol AM, Gross RW, Williams LT: **Residues in chaperonin GroEL required for polypeptide binding and release.** *Nature* 1994, **371**:614–619.

123. Camacho CJ, Thirumalai D: **Minimum energy compact structures of random sequences of heteropolymers.** *Phys Rev Lett* 1993, **71**:2505–2508.
The minimum energy maximally compact structures are enumerated for short hydrophobic/polar sequences on the square lattice. It is estimated that their number does not grow with the sequence length. On this basis, it is suggested that the rate-limiting stage of protein folding, which occurs after the rapid collapse to a random compact state, is restricted to the sampling of such maximally compact conformations.

124. Dinner A, Šali A, Karplus M, Shakhnovich E: **Phase diagram of a model protein derived by exhaustive enumeration of the conformations.** *J Chem Phys* 1994, **101**:1444–1451.
The entire ensemble for short 16-mer chains on a square lattice is enumerated. The ensemble is used to calculate rigorously the phase diagram spanned by temperature and average attraction. The results agree well with analytic heteropolymer theory.

125. Di Iorio EE, Yu W, Calonder C, Winterhalter KH, De Sanctis G, Falcioni G, Ascoli F, Giardina B, Brunori M: **Protein dynamics in minimyoglobin: is the central core of myoglobin the conformational domain?** *Proc Natl Acad Sci USA* 1993, **90**:2025–2029.

126. Stetter KO: **Life at the upper temperature border.** In *Frontiers of Life.* Edited by Tran Thanh Van JK, Mounolou JC, Schneider J, McKay C. Gif-sur-Yvette: Editions Frontieres; 1992:195–212.

127. Dill KA, Fiebig KM, Chan HS: **Cooperativity in protein-folding kinetics.** *Proc Natl Acad Sci USA* 1993, **90**:1942–1946.
The hydrophobic zipper mechanism of protein folding cooperativity is described and applied to finding the native structures of several proteins with known three-dimensional structure. Although correct native structures are not found, the best configurations have very low energies. It is also shown that real proteins have the type of topological locality predicted by the theory.

128. Shakhnovich EI, Gutin AM: **Engineering of stable and fast-folding sequences of model proteins.** *Proc Natl Acad Sci USA* 1993, **90**:7195–7199.
A novel Monte Carlo method for designing fast folding sequences is described. The method is based on the finding that a pronounced global energy minimum is conducive to rapid folding. The method is supported by statistical mechanical analysis. It is tested with hydrophobic/polar 27-mers on a cubic lattice.

129. Shakhnovich EI, Gutin AM: **A new approach to the design of stable proteins.** *Protein Eng* 1993, **6**:793–800.
The same design method as in [128] is applied to predicting sequences of four native structures corresponding to real proteins. The contact energies of Miyazawa and Jernigan are used. Statistically significant homology with natural sequences is found.

130. Ramanathan S, Shakhnovich E: **Statistical mechanics of proteins with "evolutionary selected" sequences.** *Physical Review E* 1994, **50**:1303–1312.
This is a comprehensive treatment of the statistical mechanics of designed protein sequences. A mean-field replica calculation is used to obtain the phase diagram of the system. It is also shown that 'over-design' leads to slower folding.

131. Yue K, Dill KA: **Sequence-structure relationships in proteins and copolymers.** *Phys Rev E* 1993, **48**:2267–2278.
A powerful search method for finding global energy minima of hydrophobic/polar sequences on a cubic lattice is described and applied to several sequences up to 36 monomers long. The method is based on an exhaustive search of a 'tree' that is pruned by constraints obtained from the equations relating a monomer sequence to certain aspects of the native strucuture.

132. Yue K, Dill KA: **The forces of tertiary structural organization in globular proteins.** *Proc Natl Acad Sci USA* 1995, in press.
The search method for finding global energy minima of hydrophobic/polar sequences on a cubic lattice is improved and shown to work with sequences as long as 88 residues. It is found that structures with a low degree of degeneracy tend to be highly symmetrical, similar to real proteins.

133. Thomas PD, Dill KA: **Local and nonlocal interactions in globular proteins and mechanisms of alcohol denaturation.** *Protein Sci* 1993, **2**:2050–2065.
Thermodynamic stabilities of native and denatured states of hydrophobic/polar sequences on the square lattice are estimated as a function of the strengths of local and non-local interactions. It is suggested that the alcohol denaturation of proteins, which is a transition to a helical state, is caused mainly by weakening non-local interactions and slightly enhancing local interactions.

134. Lattman EE, Fiebig KM, Dill KA: **Modeling compact denatured states of proteins.** *Biochemistry* 1994, **33**:6158–6166.
Using a cubic lattice model of an hydrophobic/polar sequence, denatured states are modelled and the results are compared with low-angle X-ray scattering data. The authors propose that denatured states consist of broad ensembles of chain backbone conformations that involve common localized hydrophobic clustering and helical contacts, which depend on the amino acid sequence.

135. O'Toole EM, Panagiotopoulos AZ: **Effect of sequence and intermolecular interactions on the number and nature of low-energy states for simple model proteins.** *J Chem Phys* 1993, **98**:3185–3190.
Energy spectra of 200 hydrophobic/polar 24-mer sequences are determined with a Monte Carlo algorithm. The low-energy states are highly degenerate. When attraction is introduced between the polar residues, the low energy conformations became significantly less degenerate.

136. Garrett DG, Kastella K: **New results on protein folding from simulated annealing.** *J Am Chem Soc* 1992, **114**:6555–6556.

137. Bromberg S, Dill KA: **Side-chain entropy and packing in proteins.** *Protein Sci* 1994, **3**:997–1009.
A lattice model with side chains is compared with a lattice model without side chains to address the role of side-chain packing in protein stability and structure. It is shown that side chain degrees of freedom oppose folding and it is proposed that packing in proteins is more like the packing of nuts and bolts in a jar than like pairwise matching of jigsaw puzzle pieces.

138. Socci ND, Bialek WS, Onuchic JN: **Properties and origins of protein secondary structure.** *Phys Rev E* 1994, **49**:3440–3443.
Using a bead model of a protein chain in continuous space, the importance of compact volume and local interactions for the formation of secondary structure is explored. It appears that compacting the chain is not enough to generate secondary structure.

139. Chan HS, Dill KA: **Intrachain loops in polymers: effects of excluded volume.** *J Chem Phys* 1989, **90**:492–509.

140. Chan HS, Dill KA: **Polymer principles in protein structure and stability.** *Annu Rev Biophys Biophys Chem* 1991, **20**:447–490.

141. Yee DP, Chan HS, Havel TF, Dill KA: **Does compactness induce secondary structure in proteins? A study of poly-ala-**

nine chains computed by distance geometry. *J Mol Biol* 1994, **241**:557–573.
Distance geometry derived poly-alanine chains are used to determine that compactness leads to considerable stabilization of secondary structure elements and that the absolute amount of secondary structure depends strongly on the criteria used to define it.

142. Skolnick J, Kolinski A: **Simulations of the folding of a globular protein.** *Science* 1990, **250**:1121–1125.

143. Godzik A, Kolinski A, Skolnick J: **Lattice representations of globular proteins: how good are they?** *J Comp Chem* 1993, **14**:1194–1202.
This is a critical evaluation of the ability of various lattices to describe real protein geometries. The most typical errors arising in low coordination number lattices are presented.

144. Kolinski A, Skolnick J: **Monte Carlo simulations of protein folding. II. Application to protein A, ROP, and crambin.** *Proteins* 1994, **18**:353–366.
This is an application of the latest version of the authors' lattice Monte Carlo folding and structure prediction method, which includes representation of side chains. Folding pathways of three proteins are described. They involve a collection of early intermediates that are followed by the rate-determining transition from compact intermediates closely resembling the molten globule state to the native-like state. The best predictions have an α-carbon root mean square error of 2.25 Å.

145. Godzik A, Kolinski A, Skolnick J: **Denovo and inverse folding predictions of protein structure and dynamics.** *J Comput Aided Mol Des* 1993, **7**:397–438.
Lattice Monte Carlo methods for protein structure prediction are reviewed.

146. Rey A, Skolnick J: **Computer modeling and folding of four-helix bundles.** *Proteins* 1993, **16**:8–28.
This lattice Monte Carlo study explores the relative importance of secondary versus tertiary interactions for the formation of a unique native state. A balance between the two is required for successful folding. A molten globule state is observed that leads to the native state by a highly cooperative freezing of the secondary structure and of the side-chain conformations.

147. Covell DG: **Folding protein α-carbon chains into compact forms by Monte Carlo methods.** *Proteins* 1992, **14**:409–420.
A bead model of a protein on a three-dimensional cubic lattice is applied to protein structure prediction for eight small proteins. An effective potential for residue–residue contacts was estimated from a database of known structures.

148. Hao M-H, Scheraga HA: **Monte Carlo simulation of a first-order transition for protein folding.** *J Phys Chem* 1994, **98**:4940–4948.
The 210 (diamond) lattice and an entropy sampling Monte Carlo method are used to simulate folding of a 38-residue polypeptide with an hydrophobic potential that included both local and global interactions. Energy is used as the reaction coordinate and a first-order transition between the denatured and native states is found. It is determined that the first-order transition arises because the entropy increases more slowly than the energy when the protein initially unfolds from the native state.

149. Hao M-H, Pincus MR, Rackovsky S, Scheraga HA: **Unfolding and refolding of the native structure of bovine pancreatic trypsin inhibitor studied by computer simulation.** *Biochemistry* 1993, **32**:9614–9631.
An extra term is added to the energy function to simulate unfolding and refolding of BPTI.

150. Chelvanayagam G, Reich Z, Bringas R, Argos P: **Prediction of protein folding pathways.** *J Mol Biol* 1992, **227**:901–916. [Published erratum appears in *J Mol Biol* 1993, **230**:695.]

151. Chelvanayagam G, Argos P: **Prediction of protein folding pathways — bovine pancreatic trypsin inhibitor.** *Cytotechnology* 1993, **11**:67–71.
A derivation of folding pathways by a greedy hydrophobic burial process is described.

152. Moult J, Unger R: **An analysis of protein folding pathways.** *Biochemistry* 1991, **30**:3816–3824.

153. Rooman MJ, Kocher J-PA, Wodak SJ: **Extracting information on folding from the amino acid sequence: accurate predictions for protein regions with preferred conformation in the absence of tertiary conformation.** *Biochemistry* 1992, **31**:10226–10238.

154. Khorasanizadeh S, Peters ID, Butt T, Roder, H: **Folding and stability of a tryptophan-containing mutant of ubiquitin.** *Biochemistry* 1993, **32**:7054–7063.

155. Kuszewski J, Clore GM, Gronenborg AM: **Fast folding of a prototypic polypeptide: the immunoglobulin binding domain of streptococcal protein G.** *Protein Science* 1994, **3**:1945–1952.
An interesting quenched flow D–H exchange experiment indicated that a collapsed state with significant NH protection is formed in the deadtime of the experiment. The authors state this in accord with a three stage folding mechanism and cite [44,46,95]. The mechanism described in detail by Kuszewski *et al.* is the 3SRS mechanism [44], which differs significantly from those in [46,95], as made clear in the main text.

156. Karplus M, Caflisch A, Šali A, Shakhnovich E: **Protein dynamics: from the native to the unfolded state and back again.** In *Modelling of bimolecular structures and mechanisms.* Edited by Pullman A *et al.*. Dordrecht: Kluwer Academic; 1995: 69–84.

M Karplus, Department of Chemistry, Harvard University, Cambridge, MA 02138, USA (permanent address).
A Šali, The Rockefeller University, 1230 York Ave, New York, NY 10021, USA.