Deciphering the timescales and mechanisms of protein folding using minimal off-lattice models

D Thirumalai* and DK Klimov

Considerable insights into the mechanisms and timescales of protein folding have been obtained from detailed studies of minimal off-lattice models. These models are coarse-grained representations of polypeptide chains. Many novel predictions of the mechanisms and timescales of the folding of proteins have been made using simulations of off-lattice models. The concepts derived from these simulations have been used to analyze the recent experiments and simulations of proteins and peptides.

Addresses
Department of Chemistry and Biochemistry, University of Maryland, College Park, MD 20742, USA
*e-mail: thirum@glue.umd.edu

Current Opinion in Structural Biology 1999, 9:197–207

http://biomednet.com/elecref/0959440X00900197

Abbreviations

ADS	atomic detailed simulations
CBA	competing basin of attraction
FCC	face centered cubic
GHT	Go-Honeycutt-Thirumalai
HT	Honeycutt–Thirumalai
I	intermediate
KPM	kinetic partitioning mechanism
MFN	multiple folding nuclei
NBA	native basin of attraction
NC	nucleation-collapse
SFN	specific folding nucleus
U	unfolded

Introduction

What are the folding routes navigated by an ensemble of denatured states of a polypeptide chain en route to the native conformation? The quest to answer this question has led to substantial advances in theory [1,2••,3,4•,5,6,7••], experiments [8–12,13•,14••] and computational strategies [15–21,22•,23•,24–26,27••]. These tools have been used to unravel the diverse mechanisms by which proteins fold. Through fast-folding experiments, it has become possible to watch folding events on submicrosecond timescales [13•,28,29••,30–34]. The principles of protein engineering [14••] are being used to successfully assess the nature of the transition-state ensembles of small proteins [35,36,37••,38••]. From a theoretical perspective, much progress has been made using minimal model representations of proteins [1]. With very few exceptions [39], most of the attention has been given to lessons from Monte Carlo studies of low-resolution lattice models [1,3,6,40–44]. Such models have been useful in testing certain general principles of the kinetics of protein folding [1,2••,41,44]. The most significant advantage of minimal models is that precisely formulated questions can be quantitatively answered using converged computations. The severe limitations of lattice models, such as the discrete representation of spatial coordinates and energies, however, prevent them from providing a realistic picture of folding pathways. It is, perhaps, for this reason that there has been a resurgence in using minimal off-lattice models to investigate various aspects of protein folding [25,26,27••,45–49]. These models offer a realistic representation of the secondary and tertiary structures, while still retaining a coarse-grained description of the polypeptide chain. This enables one to completely characterize the thermodynamics and kinetics of folding.

Both fully atomic detailed simulations (ADS) in aqueous solution and coarse-grained representations of polypeptide chains are off-lattice models. The two differ substantially in their level of detail. In the latter, all of the details on the 4–6 Å scale are averaged out and are included in the effective interactions between the beads representing the various amino acids. A full understanding of the pathways and barriers to folding requires ADS in an explicit water environment. Although such simulations of peptides [50•] and a small protein [51••] have been reported recently [52–55], they are not yet routine. Therefore, studies using minimal off-lattice models have been used to provide a conceptual framework for understanding folding mechanisms in proteins. The idea that simple representations of polypeptide chains (one or two beads per residue) could be useful has been already used by Levitt and Warshel [56] in their pioneering attempts to predict the fold of the bovine pancreatic trypsin inhibitor. The utility of such simple models for probing folding kinetics and mechanisms was demonstrated sometime ago by Honeycutt and Thirumalai [15] and by Rey and Skolnick [16].

Following these early studies, there have been several recent developments in using minimal off-lattice models to understand many aspects of protein folding [22•,23•, 24–26,27••,45,47,49]. These models have been used to address a number of important aspects of folding kinetics. They include predictions of timescales for the formation of native structures [25,27••,45], variations in the folding mechanisms in terms of the characteristics of the sequences and topologies [25,27••,49] and the clarification of the nature of the transition states of fast-folding proteins [26,57••,58,59]. Here, we provide a unified picture of the folding mechanisms for a class of proteins by synthesizing various concepts that have originated from detailed computations of minimal models, as well as ideas from polymer science.

Characteristic temperatures and phases of a polypeptide chain

Proteins are finite systems that exist in solution (the solvent of interest is water). The finiteness of such systems,

together with their marginal stability, makes the characterization of the 'phases' of proteins difficult. Nevertheless, general expectations of the putative phases of a monomeric protein can be made in analogy with polymers [7••,60]. Such an analogy would suggest that, at a sufficiently high temperature (k_BT is greater than the attractive interactions that confer a globular nature to folded proteins), the polypeptide chain is expected to display the characteristics of a random coil. This is the unfolded (U) phase of the protein. In practice (especially when the unfolded state is realized by adding excess chemical denaturants), there could be remnants of secondary structural elements present, even in random coil conformations [61]. As the temperature is lowered to the collapse temperature T_θ, we expect the transition of the chain to a compact phase. This transition is analogous to the collapse process in polymer systems. The transition at T_θ is either first or second order, depending on the nature of the interactions and the conformational fluctuations. In a protein, there are additional energy scales (due to the diversity of amino acid residues) that render a few of the exponentially large number of conformations lower in free energy than the rest. As a result, when the temperature is lowered to the folding transition temperature T_F (below T_θ), a transition to the folded native structure takes place. These general arguments suggest that, generically, we expect three phases for a protein as the temperature (or the concentration of the denaturant) is varied. They are the unfolded (U) states, an intermediate (I) compact phase, often referred to as the molten globule [62], and the native state.

The identification of these phases requires appropriate order parameters. A natural order parameter that can distinguish between U and I would be the monomer density, $\rho = N/R_G^3$, where N is the number of amino acid residues and R_G is the radius of gyration. We expect $R_G \simeq a_U N^{3/5}$ and $R_G \simeq a_I N^{1/3}$, where a_U and a_I are the effective persistence lengths of the polypeptide chain in the two phases. It follows that $\rho \approx 0$ in U and $\rho \sim O(1)$ in I for large enough N. The finite size of polypeptide chains, as well as the plausible residual structures in the U phase, however, makes ρ difficult to use in practice. The temperature dependence of R_G (or, equivalently, the specific heat) has been used to compute T_θ [15,25]. The temperature dependence of R_G in ribonuclease A [61] is similar to that seen in off-lattice models, which implies that such measurements can be used to estimate T_θ. The fluctuations in the overlap function χ (a quantity that measures how similar a given conformation is to the native state [63]) or the probability of being in the native state may be used to compute T_F [21,27••,41,44,64]. Thus, the characterization of the expected phases requires two order parameters, which implies that a coarse-grained free energy for proteins is a function of at least two variables. There are other subtle orderings that are distinguishable within a large folded protein. This could be 'nematic ordering' due to the alignment of helices [65]. The determination of such packing requires experimental probes that are currently not being employed in the protein folding area.

As a result of the finiteness and the marginal stability of the folded states of proteins, it is more useful to discuss the 'phases' of proteins in terms of the basins of attraction corresponding to the three states [5,27••,47]. This is most easily illustrated using two-state folding. Here, the relevant phases are the basin of attraction corresponding to U and the native basin of attraction (NBA). The boundary between the two is obtained by noting that the ensemble of conformations with χ less than $\chi(T_F)$ belong to the NBA [47]. This definition has been used to compute the probability of being in the NBA, $P_{NBA}(T)$ — the midpoint of which gives T_F [44,47,64]. For sequences that have an I phase, there are other basins of attraction, in which the average value of the overlap is between the U phase and in the NBA.

An important point to remember is that the stability of the phases can be altered by changing the external conditions, such as pH or urea concentration. For example, the intermediate state of apomyoglobin is more stable at pH 4 than either the native state or the unfolded state [66]. In terms of the characteristic temperatures, it has been argued that the putative phase between T_θ and T_F would correspond to the equilibrium I phase [63,67,68].

Folding rates are determined by underlying thermodynamic transitions

Since the pioneering experiments of Anfinsen [69], it has been appreciated that protein folding is a self-assembly process, that is, the information needed to specify the topology of the native state is contained within the primary sequence itself. Unlike other self-assembled systems (like micelles), the topology of the native state is relatively unique. The conformations that belong to the NBA can very rapidly interconvert among themselves on microscopic timescales. A key question that has attracted considerable attention over the past several years is what are the intrinsic, sequence-dependent properties that determine the dual requirements of the marginal stability, as well as the kinetic accessibility of the folded states of globular proteins? In other words, is there a link between folding speeds and the underlying thermodynamic properties? It has been shown, using lattice and off-lattice models, that many global aspects of the folding kinetics of proteins and RNA can be understood in terms of the characteristic temperatures T_θ and T_F [5,7••,27••,41,44,63,70]. In particular, the key factor that determines the foldability of sequences is the single dimensionless parameter:

$$\sigma = \frac{T_\theta - T_F}{T_\theta} \qquad (1)$$

which indicates how far the collapse transition temperature is from the folding temperature [41,63].

Using a variety of models, we have shown that, under conditions in which the NBA has the largest occupation

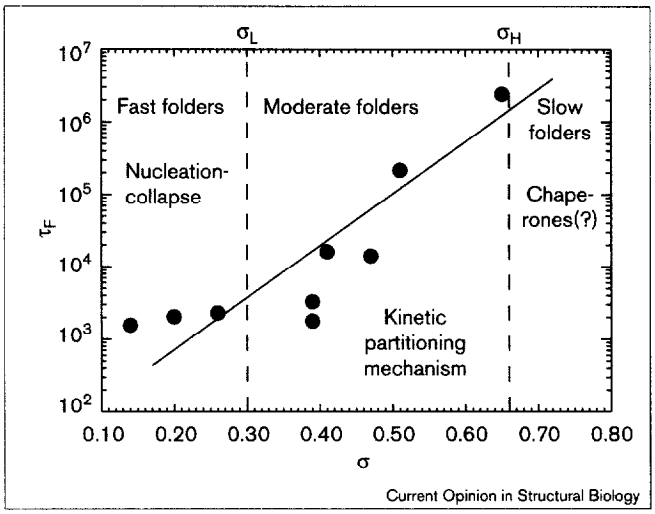
probability, the folding rates are determined by σ [27••]. In Figure 1, we show the folding times for nine sequences, each with a β turn in the native state. Remarkably, when σ changes by a factor of three or four, the folding time can increase by nearly five orders of magnitude. Such correlations have also been demonstrated in lattice models of proteins [41,44,70,71]. These studies clearly show that sequences with relatively small σ values fold rapidly under a large range of external conditions. These observations lead to a clearly stated foldability principle that states that optimal folding occurs whenever $T_F \simeq T_\theta$ [7••].

The relationship between σ and the spectrum of states and the entropy of the misfolded structures suggests that the folding rates should show a significant correlation with the Z-score, which is the difference between the energy of the native state and the average energy of the misfolded structures divided by the dispersion in the energies of the misfolded structures. This is indeed the case ([44]; Dinner *et al.*, personal communication). In contrast, there is no useful correlation between the rates of folding and the much ballyhooed energy gap (however it is defined) [41,44]. More importantly, the energy gap is not easily computable for off-lattice models [24,72].

Three recent studies using off-lattice models have confirmed that sequences with small σ values fold rapidly [22•,27••,49]. Two of these [22•,49] employ the Honeycutt–Thirumalai (HT) model [15], which has been demonstrated to be a moderate folder [25], as the starting point. Shea *et al.* [22•] constructed a number of sequences, starting with the HT sequence, by decreasing the strength of the non-native hydrophobic interactions. They found that σ becomes smaller as the strength of the non-native hydrophobic interactions is reduced. For the extreme case of zero non-native interactions, the value of σ is nearly zero. Our theory [60,63] would suggest that these sequences would fold rapidly. Nymeyer *et al.* [49] considered the Go version of the HT model (GHT) by setting all the non-native interactions to zero. They found that, for the resulting GHT model, σ is almost zero (the peak in the specific heat [T_θ] nearly coincides with the temperature T_F at which the fraction of native contacts is 0.5, see Figure 5 of [49]). Not surprisingly, the GHT sequence folds more rapidly than the HT sequence under similar folding conditions, thus confirming the validity of σ as a determinant of the folding rates. The studies of Veitshans *et al.* [27••] also showed that, by introducing diversity in the hydrophobic interactions, the value of σ can be systematically lowered. This, of course, translates into rapidly folding sequences, as shown in Figure 1.

It is clear from these studies that one can design fast folding sequences by lowering σ. Although T_F can be made to approach T_θ using an optimal design procedure, it is almost never exactly zero for protein-like sequences. We have shown previously that, in sequences with small σ values, there is compatibility between interactions on small length scales, which are responsible for secondary structures, and

Figure 1



The dependence of the folding time (τ_F) as a function of $\sigma = (T_\theta - T_F)/T_\theta$, where T_θ and T_F are the collapse and folding transition temperatures, respectively. The timescales were computed using Langevin simulations for nine sequences whose native state is a β turn. The vertical dashed lines represent approximately the boundaries in the σ axis at which the folding mechanism changes. The location of the boundaries is for guidance only. For example, between σ_L and σ_H, the polypeptide chain is expected to fold by the KPM (see Figure 2). When σ exceeds σ_H, spontaneous folding occurs with very little yield of the native state. Such sequences may require chaperones to reach the native state.

interactions on longer length scales, which confer a globular nature and specific native-state topology [7••,68]. Thus, sequences with small σ are a mathematical manifestation of the consistency principle proposed by Go sometime ago [73]. As a result of chain connectivity, there is always residual topological frustration (see below) in proteins, even when $\sigma \approx 0$, which implies that there is no ideal protein sequence with a perfect balance between local and nonlocal interactions. It is for this reason that one does not expect completely barrier-free folding in proteins. When $T_F \simeq T_\theta$, the barriers separating the folded and unfolded states are minimized, but not eliminated.

Folding mechanisms

The diversity in folding mechanisms can also be classified naturally in terms of differing values of σ. From the apparent variations in the folding rates with increasing σ, we expect that the folding mechanisms are also different. The various scenarios that arise as σ is altered may be classified as described below [27••,41].

Small σ and the nucleation-collapse mechanism

Fast folding sequences have σ values that are less than σ_L. Although we have not been able to estimate σ_L, we find that it is generally less than about a quarter. These sequences display two-state behavior, both thermodynamically and kinetically [27••,41,44,64,74•]. The thermodynamic transition at T_F is first order and cooperative [3,4•]. As a result of the finite size of polypeptide

chains, the transition at T_F is not infinitely cooperative. Recently, we have introduced a dimensionless measure of cooperativity [64]. This index measures the strength of the transition at T_F. By analyzing equilibrium data on a number of proteins, we showed that, when unfolding is induced by chemical denaturants, the cooperativity index is in the range 5–15 [64], whereas the corresponding values would be nearly infinite in an extremely cooperative transition (the crystallization of liquids for example) [75]. For sequences with small σ, the native state is reached in a single kinetic step, that is, typically, one observes first-order kinetics in the transition from U to the NBA [27••]. These characteristics of two-state folders, which have been seen in experiments (e.g. [8,76]) and in minimal models of proteins [22•,27••,44,49], raise fundamental questions: what is the timescale for the transition to the native state starting from an ensemble of denatured states and what are the pathways, mechanisms and transition states that are involved en route to the native state?

Timescales

The timescales for forming fragments of structures, such as β turns and α helices, have been obtained using Langevin simulations of simple off-lattice models [27••,45,57••]. It is predicted that α helices form in 10–100 ns, depending on the temperature [45]. These values are in good accord with experiments [29••,33], as well as other theoretical estimates. The rates of β-turn formation are currently being investigated in detail, both computationally [27••,57••] and by fast-folding experiments [29••]. Using a number of sequences that have a β turn as the native state, we have shown that they can fold in about 10 μs [27••,57••]. Not surprisingly, these theoretical predictions were subsequently confirmed by Munoz *et al.* [29••], who showed that a peptide (a fragment of the B1 domain of the immunoglobulin-binding protein [GB1]) forms a β turn in 6 μs.

A tentative expression for the timescales for accessing the NBA for two-state folders has been given using arguments from polymer theory [60]. It has been argued that optimized sequences, which fold in two-state manner and reach the NBA under favorable conditions, fold in times given by:

$$\tau_F \approx \frac{\eta a}{\gamma} f(\sigma) N^{\omega} \qquad (2)$$

where ω is between 3.8 and 4.2, η is the viscosity, a is an effective persistence length, γ is the average surface tension and N is the number of amino acids in the protein. In general, $f(\sigma)$ depends exponentially on σ (see Figure 1). We emphasize that this prediction gives estimates that are accurate only to within an order of magnitude.

It is instructive to estimate the folding times for a few two-state folders using Equation 2. For this purpose, we assume that $f(\sigma)$ is on the order of unity because σ is expected to be small. The folding time for the two-state folding peptide investigated by Munoz *et al.* [29••] is predicted to be in the range of 4–7 μs using Equation 2, which agrees extremely well with measurements [29••], as well as with Langevin simulations [57••]. In obtaining this absolute time, we have used the fact that the prefactor, $\eta a/\gamma$, is about 0.1 ns, which is obtained using typical values for η, a and γ [60]. We also estimate that the folding time for the much studied chymotrypsin inhibitor 2 (CI2) is in the range of 1–4 ms, whereas the 67 residue protein cspB is predicted to fold in range of 1–5 ms. The experimental folding times for CI2 and cspB are 18 ms [35] and 1.5 ms [77], respectively. Our estimates of folding times using Equation 2 agree extremely well with experiments [35,77]. Thus, a combination of the theoretical arguments given in [60] and Langevin simulations of minimal off-lattice [45,57••] may suffice to obtain estimates of the overall timescales of the folding of small proteins.

Sequences with small σ fold in such a way that the process of collapse and the acquisition of the native conformation are almost simultaneous. This would imply that $\tau_c/\tau_F \sim 0(1)$, where τ_c is the time for specific collapse to occur [60]. Recent experiments on protein L (D Baker, personal communication) confirm our theoretical prediction. It is also consistent with the notion that, in the nucleation-collapse (NC) mechanism, it is not advantageous to have preformed native contacts in the denatured states [14••]. Therefore, the efficient folding of two-state folders reflects the fact that collapse and folding are almost synchronous.

Nucleation-collapse mechanism and transition states

It is natural to postulate that because, at T_F, a first-order folding transition takes place, a nucleation-like mechanism is operative in two-state folders. This was suggested by Wetlaufer [78] a long time ago to explain the efficient folding of proteins. For sequences with small σ, we have suggested that the acquisition of the native state (by searching for one of the critical nuclei) and the collapse processes are almost simultaneous [7••,60]. Thus, this process is referred to as the NC mechanism. In recent years, there has been renewed interest in understanding the nature of the NC mechanism using protein engineering experiments [35] and simulations of lattice and off-lattice models [25,26,58,74•,79].

The average size of the critical nuclei

When a liquid, say argon, is slowly cooled below its melting temperature, it crystallizes into a face centered cubic (FCC) lattice. Kinetically, this process is triggered by the formation of droplets that have crystalline (not necessarily FCC) features. There is a distribution in the sizes of these droplets [75]. If the radius of the droplet exceeds a critical value (a nucleus), then the FCC crystal is realized with overwhelming probability. The analogy to crystallization has been used to predict the average size of the critical

nucleus in single-domain proteins. By balancing the driving force towards the native state and the surface tension of the 'droplet' containing the residues in the nucleus, Guo and Thirumalai [25] showed that the average size of the critical nucleus is between 10–15 residues. This is in accord with the estimate made by Matheson and Scheraga [80] some time ago using a very different procedure. By refining an earlier theory, Wolynes [81] suggests that roughly a third of the residues constitute the critical nucleus. A more realistic treatment (JD Bryngelson, D Thirumalai, unpublished data), using experimentally measured quantities, also shows that the average size of the nucleus is between 10–15 residues. Thus, a variety of theoretical approaches leads to consistent results for the average size of the nucleus for single-domain proteins.

The droplet picture of Guo and Thirumalai [25], as well as that of Wolynes [81], shows that, in general, the width of the transition region is relatively broad. Both the topology of the native state and the external conditions determine the transition state width and the heterogeneity of the transition state structures [4•,59]. The relatively broad transition region is suggestive of a plastic transition-state ensemble [26].

The multiple folding nuclei model for nucleation-collapse mechanism

The relatively small size of the nuclei implies that only a subset of residues is in contact in the putative transition states. The structural characteristics of the conformations that constitute the critical nuclei have been investigated using both lattice and off-lattice models [26,58]. Based on an extensive analysis of the HT model, which has a β barrel as the native state, Guo and Thirumalai [26] showed that there are several nucleation regions. The contacts between the residues in the nucleation regions occur with varying probability in the transition state. It was further demonstrated that the most probable nucleation regions occur close to the turn region, which is in accord with experiments [37••,38••]. The formation of such nucleation regions is responsible for the subsequent rapid assembly of the native conformation.

When averaged over the denatured ensemble, some of the contacts (those close to the loop regions in the β-sheet models) occur with a higher probability than others. This observation allowed us to propose the multiple folding nuclei (MFN) model, which asserts that, in the folding nuclei, there is a distribution of contacts, with some occurring with higher probability than others [74•,79]. The structures corresponding to the critical nuclei are topology preserving, that is, the nucleation regions have native-like structures. As a result of the nature of the topology-preserving transition-state structures, the number of structures corresponding to the critical nuclei is not exponentially large. In the HT four-turn β-barrel models, contacts between beads near the middle two turns occur with a higher probability than others. These high probability contacts are formed as a result of the flexibility of loops. The importance of the entropy associated with loop formation in the transition states was stressed by Guo and Thirumalai [25] and also has been recently pointed out by Pande *et al.* [4•].

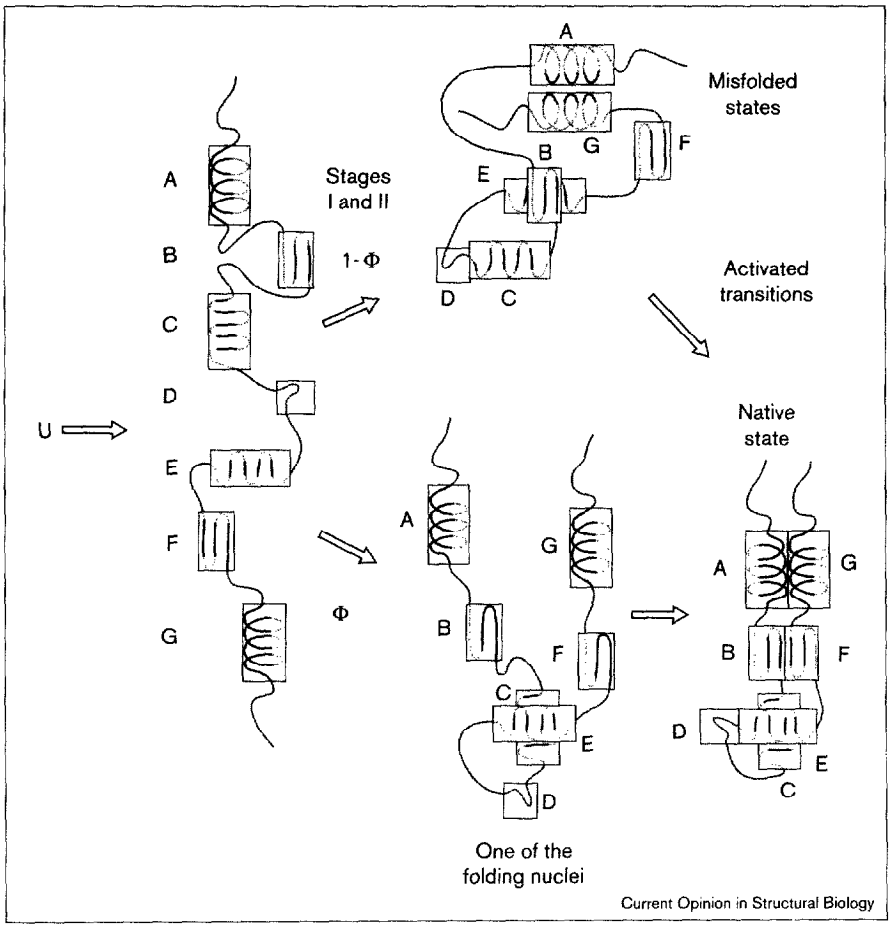
The validity of the MFN was further illustrated by a detailed analysis of lattice models of proteins [74•]. The probability distribution of the size and number of residues and other structural characteristics unambiguously demonstrated that there are many nuclei involved in the folding process. On an average, there are some contacts that are found with a higher probability than others. The simultaneous participation of a subset of native contacts in each and every member of the transition-state ensemble is not observed. If there is a subset of contacts that participate with unit probability, as was postulated in the specific nucleus model (SFN) [58], then the distribution of contacts would have a peak at unit probability. The absence of such a peak in the distribution functions was used to show that the SFN model is not a valid description of the nucleation-collapse mechanism, at least for simple models [74•,79].

A different picture of the folding nuclei is advocated by Wolynes and co-workers [59]. He argues that different native contacts participate in the transition state with different probabilities. The implication of this is that, in the transition state, all the contacts are involved, with some participating with a higher probability than others, in every molecule en route to the native state. This, of course, leads to a multiple, delocalized (involving the entire protein) nucleus. If we assume that only the very high probability contacts are important, then this model is equivalent to the MFN model. It should be emphasized that both the MFN model and the delocalized nuclei picture are dramatically different from the SFN model.

Topological frustration, intermediate σ values, and the kinetic partitioning mechanism

Sequences with intermediate values of σ are moderate folders and their folding kinetics is considerably more complicated. The qualitative aspects of the kinetics for these sequences, which invariably involve intermediates, can be understood in terms of topological frustration [7••]. A protein sequence contains about 55% hydrophobic residues. The linear density of hydrophobic residues along the polypeptide chain is roughly constant, implying that the hydrophobic residues are spread throughout the chain. As a result, on any length scale greater than the persistance length (l), there is a propensity for certain residues to form tertiary contacts under folding conditions. The structures, which arise as a result of contacts between residues that are in close proximity, would be in conflict with the global native fold. The incompatibility of structures on local scales with the near unique native structure leads to topological frustration. It is important to realize that topological frustration is inherent to all foldable sequences and is a direct consequence of the polymeric nature of proteins, as well as of the competing interactions (hydrophobic residues, which prefer the formation of compact structures, and hydrophilic residues, which are better accommodated

Figure 2



A sketch of the KPM. The letters indicate blocks of local secondary structural elements in a protein sequence. A fraction of the molecules (Φ) reaches the native conformation rapidly. The transition state corresponding to one of the folding nuclei is displayed. It is native-like, in that a subset of native contacts is formed. The folding nuclei are an expanded form of the native state. The remaining fraction (1−Φ) reaches the native state by three-stage kinetics. Stages I and II lead the polypeptide chain to one of the CBAs (misfolded states). The mechanisms underlying these stages are described in the text. The last stage consists of activated transitions from the CBAs to the native state. The partition factor Φ can be altered by changing the external conditions. The KPM mechanism has been mapped out in some detail in the refolding experiments on lysozyme.

by extended conformations). It is for this reason that an ideal protein, which has complete compatibility between local and nonlocal interactions, does not exist, as was first recognized by Go [73]. Thus, topological frustration is an inherent property of all polypeptide chains. Even the Go model is topologically frustrated. A consequence of topological frustration is that the underlying energy landscape is rugged, consisting of many minima separated by barriers of varying heights.

The nature of the low lying minima in the rugged energy landscape can be qualitatively described [5,82]. On the length scale l, there are many ways of forming structures that are in conflict with the global fold. It is expected that most of these structures would have high free energies and, hence, would be unstable to thermal fluctuations. We expect a certain number of these structures (predicted to grow only as lnN [82]) to have low free energies. There could be significant overlap between these structures and the native fold and, hence, these structures could be viewed as being native-like. Certain regions in these structures could be disordered or misfolded. These structures would correspond to the competing basins of attraction (CBAs), which act as kinetic traps that slow the folding process [27••].

The basic consequences of the complex free energy surface arising from topological frustration leads naturally to the kinetic partitioning mechanism (KPM) [5,7••,25,26,27••,60,83]. Imagine an ensemble of denatured molecules in search of the native conformation. This is the experimental situation that arises when the concentration of denaturant is decreased. It is clear that a fraction of the molecules, Φ (the partition factor), would reach the NBA rapidly, without being trapped in the low energy minima. The remaining fraction would be trapped in one or more of the minima and reach the NBA by activated transitions only on longer timescales. The value of Φ depends on the sequence and is determined by σ. Thus, because of topological frustration, the initial pool of molecules partitions into fast folders and slow folders. A schematic sketch of the KPM is given in Figure 2.

The qualitative aspects of KPM (described above) for moderate folding sequences were explicitly verified using the HT model by Guo and Thirumalai [25,83]. They showed that the fraction of molecules Φ that reaches the NBA rapidly does so in a two-state manner, without populating any intermediates. This pool of molecules on the fast track reaches the NBA through the NC mechanism, as

found for sequences with small σ. The only difference is that, for sequences with small σ, the amplitude Φ is nearly unity [27••]. The timescale for this fast track is approximately given by Equation 2.

A detailed kinetic analysis of the remaining fraction of molecules ($1-\Phi$) showed that they reach the NBA through a three-stage multipathway mechanism [25,27••,60,63]. According to this model, the polypeptide chain initially collapses to a compact structure. Such a collapse is nonspecific and occurs on a timescale of the order of 1 μs for a protein consisting of about 100 amino acid residues [60]. The structures that are accessed on this timescale are determined by loop formation kinetics and favorable local interactions. In the second stage (driven by energetic biases), the polypeptide chain diffusively searches for low-energy structures. As a result of the initial nonspecific collapse, the polypeptide chain reaches one of the CBAs at the end of the second stage. The final stage involves activated transitions from one of the CBAs to the NBA. There are many pathways that connect the denatured states to the CBAs, but the number of paths from the CBAs to the NBA is limited. This suggests that the major transition states for the ($1-\Phi$) fraction of molecules, which occur late in the folding process, are sparse. The approximate timescales for each of the three stages have been given [60].

Experimental validation of the kinetic partitioning mechanism in proteins and RNA

The clearest verification of the KPM comes from detailed refolding experiments on the hen-egg lysozyme by Kiefhaber and Dobson and co-workers [11,84••]. The experiments by Kiefhaber [11] were the first to show that there is a fast track for folding, in which lysozyme assembles very rapidly (on the order of tens of milliseconds). Using interrupted refolding experiments, Kiefhaber showed that about 15% of the molecules ($\Phi \sim 0.15$) reach the folded state in about 50 ms, whereas the remaining fraction of molecules reaches the native state in about 400 ms. A very detailed map of the parallel pathways in lysozyme folding has been made using a combination of hydrogen-exchange labeling experiments and the binding of a fluorescently labeled inhibitor, by Mantagne *et al* [84••]. These authors showed that about 25% ($\Phi = 0.25$ at pH 5.5 and 20°) of the molecules fold by the fast track. The timescale for the assembly of the native state by the slow process exceeds 350 ms. A theoretical estimate for the slow folding process is about 100 ms [5], which is in fair agreement with experiments. Although the overall validity of the KPM has been demonstrated in the refolding experiments on lysozyme, the nature of the transition states in the slow and fast processes has not been clarified in experiments.

Although the folding of large RNAs is not focus of this article, it is worth mentioning that RNA also refolds by the KPM [5]. Using a gel assay to monitor the folding of the *Tetrahymena* group I intron, Pan *et al.* [85] showed that the folding of large RNA involves parallel pathways. The value of Φ was found to be around 0.1. These experiments and those on proteins suggest that the KPM may be valid for the folding of proteins and RNA.

Atomic detailed simulations

A complete description of folding pathways requires ADS in aqueous solution, provided that it is assumed that interaction potentials are known reasonably accurately. As a result of severe constraints on the simulation times, ADS have not been routinely used to infer folding pathways. Until now, unfolding simulations at elevated temperatures, introduced by Daggett and Levitt [52], have been used to gain insights into the plausible nature of the transition states in the refolding process. Such methods have been duplicated by others [53,86,87]. In the past year, direct ADS have been reported for peptides [50•,55]. Very recently, the results of a long molecular dynamics simulation with explicit inclusion of solvent of a villin headpiece subdomain (HP-36), a 36-residue protein, have been published [51••,88]. These studies, especially those on HP-36, demonstrate the efficacy of MD simulations, with the current force-fields, in understanding the energy landscape dictating the folding process.

Folding of peptide fragments

Dyson, Wright and co-workers [89] have undertaken systematic NMR studies of a class of peptide fragments that form reverse type VI turns. These studies have identified several peptide sequences that have a tendency to fold (to form a turn). The stability of one of these peptides, *cis*-AYPYD, was examined using molecular dynamics simulations by Demchuck *et al.* [55], who showed that the NMR structure did not undergo global unfolding during the course of the simulations. The folding kinetics of another series of peptides (SYPFDV and their sequence variants) for which NMR structures are available was investigated by Elber and co-workers [50•]. This study, which was the first to demonstrate the reversible folding of a peptide in aqueous solution, yielded structures that are in good agreement with those determined by NMR spectroscopy. This is encouraging because it demonstrates that, at least for small protein fragments in solution, structures can be predicted using standard energy functions and novel simulation protocols. More importantly, it was shown that the folded structures can form at room temperature, starting from initial random-coil states. Elber and colleagues [50•] have predicted that the type VI turn in these peptides can form in about a nanosecond, which, in principle, can be confirmed experimentally. Surprisingly, the analysis of the pathways seen in the formation of the turn shows that several features expected in the folding of proteins (such as multistage kinetics and the presence of off-pathway processes leading to KPM) are already manifested in these small peptides. This suggests that the detailed examination of structure formation in such studies may offer a glimpse of the underlying energy landscape of proteins.

In a subsequent study, van Gunsteren and co-workers [90] have reported a series of molecular dynamics simulations (in methanol) on a non-natural β heptapeptide that adopts a helical structure in the folded state. Using a 50 ns simulation, they found that this peptide can fold reversibly in about 10 ns, starting from denatured states. Although they did not explore the folding kinetics and pathways in detail, as was carried out on the turn-forming peptides, their study confirms qualitatively the findings of Mohanty *et al.* [50•].

Refolding of proteins

A significant step forward in dissecting the folding pathways of the 36-residue villin headpiece subdomain (HP-36) has been taken by Duan and Kollman [51••], using a 1 μs trajectory generated by standard ADS at room temperature [88]. As HP-36 is expected to be a two-state folder, σ would be relatively small. If this is the case, the folding time can be theoretically estimated using Equation 2. For HP-36, we estimate a from the radius of gyration of the compact states probed in the simulation [51••]. They found that $R_G \sim 8.7$ Å $\approx aN^{1/3}$, which gives $a = 2.6$ Å, which is smaller than what is generally predicted for proteins. Using values of $\eta \approx 0.01P$ and $\gamma \approx 25$–50 cal/molÅ^2, it is found using Equation 2 that $\tau_F \approx 6$–50 μs, which is only somewhat larger than the simulation timescale [51••]. Thus, their study, in principle, can provide a picture of many of the early events in the folding of a small protein.

According to Duan and Kollman [51••], 50% of the helical content of HP-36, which has three short helices connected by a loop and a turn, is formed in about 60 ns. This estimate for helix formation time is consistent with experiments [33], as well as with estimates based on simulations of off-lattice models [45]. On a slightly longer timescale, there appears to be the formation of certain tertiary contacts, which is signaled by a reduction in the radius of gyration. This short-lived state expands at later times. This process is reminiscent of the search for folding nuclei seen in off-lattice models of a β barrel [26]. In such processes, the formation and break up of native contacts occur until the critical nuclei are found. A verification of this conjecture for HP-36 would require the generation of several independent trajectories of about 0.5 μs each.

Conclusions

The use of minimal off-lattice models, which can serve as a bridge between all-atom simulations and experiments, has given rise to a number of important predictions for both the overall timescales and the mechanisms of folding. These advances, together with new experimental measurements, have helped to sharpen our understanding of the processes involved in the assembly of proteins. In particular, the NC mechanism helps to rationalize the efficient and fast folding of many two-state folders [14••]. Although there are many unsettled issues about the NC processes, there are several points that are undisputed. For example, it is recognized that the average size of the critical nuclei is small and they consist mostly of residues that are close together in sequence space. The need for a few long-range contacts in the NC process has also been firmly established in experiments [35,37••], as well as in simulations of simplified models [26,74•]. Furthermore, the relationship between the collapse and folding processes in these two-state folders is also becoming clear, thanks to small-angle X-ray scattering experiments (D Baker, personal communication). These experiments show that, in proteins that fold efficiently, the processes of collapse and the acquisition of the native structure are nearly simultaneous, which, again, is a converging point for experiment and theory [60]. These observations lead to the optimistic conclusion that many features of two-state folding proteins are under both experimental and theoretical control.

The proteins that fold by the KPM pose greater challenges. Although the outlines of the KPM have been noted for the refolding of lysozyme [84••] (and perhaps barnase [91]), several issues require additional theoretical and experimental work. In particular, the nature of the mechanisms in the slow and fast folding tracks is not fully understood. As there appears to be an accumulation of intermediates in these proteins, it would interesting to characterize the structure and timescale of the formation of these species. In this vein, a combination of minimal off-lattice simulations of moderate folding sequences and ADS of the kind initiated for HP-36 and other smaller peptides may be useful. We expect advances in many fronts to provide answers to these and other challenging problems in the not too distant future.

Acknowledgements

This work was done while DT was in residence as a fellow in the Institute for Advanced Studies, Hebrew University of Jerusalem. DT is grateful to the members of the Institute for their hospitality. We are grateful to JD Bryngelson, K Plaxco, H Orland and W Eaton for pertinent discussions during the writing of this article. We would like to thank D Baker for allowing us to quote the results of his experiments on protein L. This work was supported by a grant from the National Science Foundation.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Dill KA, Bromberg S, Yue K, Fiebig KM, Yee DP, Thomas PD, Chan HS: **Principles of protein folding – a perspective from simple exact models.** *Protein Sci* 1995, **4**:561-602.

2. •• Dill KA, Chan HS: **From Levinthal to pathways to funnels.** *Nat Struct Biol* 1997, **4**:10-19.
An excellent review of theoretical studies of protein folding. The statistical mechanical perspective of folding is summarized and the authors present a simple description of the expected scenarios for the kinetics of folding. Comparisons with experimental results make this a handy review.

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

4. • Pande VS, Grosberg AY, Tanaka T, Rokhsar DS: **Pathways for protein folding: is a new view needed?** *Curr Opin Struct Biol* 1998, **8**:68-79.
This review analyzes the lessons learned from the simulation of simple lattice models of proteins, with emphasis on design techniques, energy landscapes

and transition-state characterization. The authors discuss different views on transition state classes derived using a suitable reaction coordinate for protein folding. An attempt is made to reconcile the classical (hierarchical) view of the folding reaction with the view point suggested by the statistical mechanical approach.

5. Thirumalai D, Woodson S: **Kinetics of folding of proteins and RNA.** *Accounts Chem Res* 1996, **29**:433-439.

6. Karplus M, Sali A: **Theoretical studies of protein folding and unfolding.** *Curr Opin Struct Biol* 1995, **5**:58-73.

7. •• Thirumalai D, Klimov DK, Woodson SA: **Kinetic partitioning mechanism as unifying theme in the folding of biomolecules.** *Theor Chem Acc* 1997, **1**:23-30.
This review shows that the kinetic partitioning mechanism (KPM) derived from the concept of topological frustration explains the diverse folding behavior observed in proteins. Illustrations from protein and RNA literature are used to provide experimental support.

8. Jackson SE, Fersht AR: **Folding of chymotrypsin inhibitor 2: 1 Evidence for a two-state transition.** *Biochemistry* 1991, **30**:10428-10435.

9. Sosnick TR, Mayne L, Hiller R, Englander SW: **The barriers in protein folding.** *Nat Struct Biol* 1994, **1**:149-156.

10. Radford SE, Dobson CM: **Insights into protein folding using physical techniques: studies of lysozyme and α-lactalbumin.** *Philos Trans R Soc Lond Biol* 1995, **348**:17-25.

11. Kiefhaber T: **Kinetic traps in lysozyme folding.** *Proc Natl Acad Sci USA* 1995, **92**:9029-9033.

12. Schindler T, Herrler M, Marahiel MA, Schmid FX: **Extremely rapid folding in the absence of intermediates.** *Nat Struct Biol* 1995, **2**:663-673.

13. • Eaton WA, Munoz V, Thompson PA, Chan CK, Hofrichter J: **Submillisecond kinetics of protein folding.** *Curr Opin Struct Biol* 1997, **7**:10-14.
A detailed account of fast elementary events (such as the assembly of α helices and β turns) in protein folding is presented using novel experimental techniques. The authors define interesting conceptual questions, such as the upper bound to the rate of folding of single-domain proteins. They argue, using theoretical ideas and experiments, that, generally, proteins cannot fold faster than 1 μs.

14. •• Fersht AR: **Nucleation mechanisms in protein folding.** *Curr Opin Struct Biol* 1997, **7**:3-9.
Protein engineering methods, pioneered by the author, are used to review evidence for the nucleation-collapse (NC) mechanism of two-state proteins. A lucid description of the NC mechanism in proteins is provided and contrasted with the well-known nucleation processes in other situations (the condensation of vapor, for example).

15. Honeycutt JD, Thirumalai D: **Metastability of the folded states of globular proteins.** *Proc Natl Acad Sci USA* 1990, **87**:3526-3529.

16. Rey A, Skolnick J: **Comparison of lattice Monte Carlo dynamics and Brownian dynamics folding pathways of α-helical hairpins.** *Chem Phys* 1991, **158**:199-219.

17. Wilson C, Doniach S: **A computer model to dynamically simulate protein folding: studies with crambin.** *Proteins* 1989, **6**:193-209.

18. Shakhnovich EI, Gutin AM: **Engineering of stable and fast folding sequences of model proteins.** *Proc Natl Acad Sci USA* 1993, **90**:7195-7199.

19. Hao MH, Scheraga HA: **Monte Carlo simulations of a first order transition for protein folding.** *J Phys Chem* 1994, **98**:4940-4948.

20. Socci ND, Onuchic JN: **Kinetic and thermodynamic analysis of protein-like heteropolymers: Monte Carlo histogram technique.** *J Chem Phys* 1995, **103**:4732-4744.

21. Guo Z, Brooks CL III, Boczko EM: **Exploring the folding free energy surface of a three-helix bundle protein.** *Proc Natl Acad Sci USA* 1997, **94**:10161-10166.

22. • Shea JE, Nochomovitz YD, Guo Z, Brooks CL III: **Exploring the space of protein folding Hamiltonians: the balance of forces in a minimalist β-barrel model.** *J Chem Phys* 1998, **109**:2895-2903.
On the basis of an off-lattice Honeycutt–Thirumalai β-barrel model, an interesting method for studying the role of native versus non-native interactions is proposed. It is shown that weakening the non-native attractive interactions leads to more cooperative folding and reduces the temperature interval between the collapse and folding temperatures, which, in turn, assures more efficient folding of proteins.

23. • He S, Scheraga HA: **Brownian dynamics simulations of protein folding.** *J Chem Phys* 1998, **108**:287-300.
This study reports the application of molecular dynamics simulations in dihedral angle space for protein folding thermodynamics and kinetics. Based on the performance of this algorithm in an off-lattice protein model, the authors conclude that it is faster and computationally more stable than traditional molecular dynamics performed in Cartesian coordinates.

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

25. Guo Z, Thirumalai D: **Kinetics of protein folding: nucleation mechanism, timescales, and pathways.** *Biopolymers* 1995, **36**:83-103.

26. Guo Z, Thirumalai D: **The nucleation-collapse mechanism in protein folding: evidence for the non-uniqueness of the folding nucleus.** *Fold Des* 1997, **2**:377-391.

27. •• Veitshans T, Klimov DK, Thirumalai D: **Protein folding kinetics: timescales, pathways, and energy landscapes in terms of sequence dependent properties.** *Fold Des* 1997, **2**:1-22.
A systematic study of a number of fast and slow folding off-lattice sequences shows that folding timescales, the nature of folding kinetics and energy landscapes are directly linked to a simple thermodynamic parameter $\sigma = (T_\theta - T_f)/T_\theta$. In particular, the authors show that, for two-state proteins ($\sigma \approx 0$), the process of collapse and the acquisition of the native conformation occur almost simultaneously. Sequences with small σ can fold efficiently while maintaining their stability over a wide range of external conditions. These results generalize the findings for lattice models [41,44,63,70].

28. Thompson PA, Eaton WA, Hofrichter J: **Laser temperature jump study of the helix-coil kinetics of an alanine peptide interpreted with a 'kinetic zipper' model.** *Biochemistry* 1997, **36**:9200-9210.

29. •• Munoz V, Thompson PA, Hofrichter J, Eaton WA: **Folding dynamics and mechanism of β-hairpin formation.** *Nature* 1997, **390**:196-199.
This paper reports the first measurement of the formation of a β hairpin using novel fast-folding experiments. The authors show that a β hairpin in a 16-mer peptide fragment of protein G forms in 6 μs. These experiments provide a direct validation of the predictions made in [27••,57••].

30. Eaton WA, Munoz V, Thompson PA, Henry ER, Hofrichter J: **Kinetics and dynamics of loops, α-helices, β-hairpins, and fast-folding proteins.** *Accounts Chem Res* 1999, **31**:745-753.

31. Bopp MA, Jia Y, Li L, Cogdell RJ, Hochstrasser RM: **Fluorescence and photobleaching dynamics of single light-harvesting complexes.** *Proc Natl Acad Sci USA* 1997, **94**:10630-10635.

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

33. Williams S, Causgrove TP, Gilmanshin R, Fang KS, Callender RH, Woodruff WH, Dyer RB: **Fast events in protein folding: helix melting and formation in a small peptide.** *Biochemistry* 1996, **35**:691-697.

34. Ladurner AG, Fersht AR: **Upper limit of the timescale for diffusion and chain collapse in chymotrypsin inhibitor 2.** *Nat Struct Biol* 1999, **6**:28-31.

35. Itzhaki LS, Otzen DE, Fersht AR: **The structure of the transition state of chymotrypsin inhibitor 2 analyzed by protein engineering methods: evidence for a nucleation-condensation mechanism for protein folding.** *J Mol Biol* 1995, **254**:260-288.

36. Viguera AR, Serrano L, Wilmanns M: **Different folding transition states may result in the same native structure.** *Nat Struct Biol* 1996, **3**:874-880.

37. •• Grancharova VP, Riddle DS, Santiago JV, Baker D: **Important role of hydrogen bonds in the structurally polarized transition state for folding of the src SH3 domain.** *Nat Struct Biol* 1998, **5**:714-720.
Using ϕ value analysis, it was concluded that part of the native structure is highly structured in the transition state of the src SH3 domain. These data are taken to distinguish folding scenarios for the src SH3 domain from those observed in chymotrypsin inhibitor 2 and the P22 Arc repressor. The role of the native topology in determining the transition-state structure is underscored.

38. •• Martinez JC, Pisabarro MT, Serrano L: **Obligatory steps in protein folding and the conformational diversity of the transition state.** *Nat Struct Biol* 1998, **5**:721-729.
By studying the folding kinetics of mutants of the α-spectrin SH3 domain, the authors suggest that the transition states of β-sheet proteins are well structured as a result of the constraints of 'mechanical nuclei'.

39. Kolinski A, Skolnick J: **Monte Carlo simulations of protein folding. I. Lattice model and interaction scheme.** *Proteins* 1994, **18**:338-352.

40. Socci ND, Onuchic JN: **Folding kinetics of protein-like heteropolymers.** *J Chem Phys* 1994, **101**:1519-1528.

41. Klimov DK, Thirumalai D: **Factors governing the foldability of proteins.** *Proteins* 1996, **26**:411-441.

42. Unger R, Moult J: **Local interactions dominate folding in a simple protein model.** *J Mol Biol* 1996, **259**:988-994.

43. Pande VS, Grosberg AY, Tanaka T: **On the theory of folding kinetics for short proteins.** *Fold Des* 1997, **2**:109-114.

44. Klimov DK, Thirumalai D: **Linking rates of folding in lattice models of proteins with underlying thermodynamic characteristics.** *J Chem Phys* 1998, **109**:4119-4125.

45. Klimov DK, Betancourt MR, Thirumalai D: **Virtual atom representation of hydrogen bonds in minimal off-lattice models of helices: effect on stability, cooperativity and kinetics.** *Fold Des* 1998, **3**:481-496.

46. Irback A, Peterson C, Potthast F, Sommelius O: **Local interactions and protein folding: a 3D off-lattice approach.** *J Chem Phys* 1997, **107**:273-290.

47. Guo Z, Brooks CL III: **Thermodynamics of protein folding: a statistical mechanical study of a small all β protein.** *Biopolymers* 1997, **42**:745-757.

48. Berry RS, Elmaci N, Rose JP, Vekhter B: **Linking topography of its potential surface with the dynamics of folding of a protein model.** *Proc Natl Acad Sci USA* 1997, **94**:9520-9524.

49. Nymeyer H, Garcia AE, Onuchic JN: **Folding funnels and frustration in off-lattice minimalist protein landscapes.** *Proc Natl Acad Sci USA* 1998, **95**:5921-5928.

50. • Mohanty D, Elber R, Thirumalai D, Beglov D, Roux B: **Kinetics of peptide folding: computer simulations of SYPFDV and peptide variants in water.** *J Mol Biol* 1997, **272**:423-442.
A thorough molecular dynamics study of a full atomic model for a short peptide reveals that its folding bears features that are characteristic of larger proteins, such as kinetic traps, kinetic partitioning and so on.

51. •• Duan Y, Kollman PA: **Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution.** *Science* 1998, **282**:740-744.
This paper represents a milestone in all-atom simulations of proteins. The authors successfully integrated Newton equations of motion for the villin headpiece subdomain (HP-36) with explicit inclusion of solvent molecules for 1 μs. The folding time for this protein is expected to be about 6–50 μs (see Equation 2), which is well within the reach of present day computations. A detailed analysis of such trajectories is expected to give considerable insight into early events in the folding of small proteins.

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

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

54. Boczko EM, Brooks CL III: **First principle calculation of the folding free energy of a three helix bundle protein.** *Science* 1995, **21**:393-396.

55. Demchuk E, Bashford D, Case DA: **Dynamics of a type VI reverse turn in a linear peptide in aqueous solution.** *Fold Des* 1997, **2**:35-46.

56. Levitt M, Warshel A: **Computer simulation of protein folding.** *Nature* 1975, **253**:694-697.

57. •• Klimov DK, Thirumalai D: **Viscosity dependence of the folding rates of proteins.** *Phys Rev Lett* 1997, **79**:317-320.
The viscosity dependence of rates of folding was examined using a number of simple models. The authors argue (especially for two-state folders) that, because the native conformation may be directly accessed from relatively open structures, the internal viscosity of the polypeptide chain is not significant in determining the folding rates. Internal viscosity will be important in the transition from misfolded semi-compact states to native state (see Figure 2) in sequences exhibiting KPM. The authors also predicted that, typically, β-turn structures can form in about 10 μs, which was subsequently verified in fast folding experiments by Munoz *et al.* [29••].

58. Shakhnovich E, Abkevich V, Ptitsyn O: **Conserved residues and the mechanism of protein folding.** *Nature* 1996, **379**:96-98.

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

60. Thirumalai D: **From minimal models to real proteins: timescales for protein folding.** *J Physique I* 1995, **5**:1457-1467.

61. Sosnick TR, Trewhella J: **Denaturated states of ribonuclease A have compact dimensions and residual secondary structure.** *Biochemistry* 1992, **31**:8329-8335.

62. Ptitsyn OB: **The molten globule state.** In *Protein Folding.* Edited by Creighton TE. New York: Freeman; 1994:243-300.

63. Camacho CJ, Thirumalai D: **Kinetics and thermodynamics of folding in model proteins.** *Proc Natl Acad Sci USA* 1993, **90**:6369-6372.

64. Klimov DK, Thirumalai D: **Cooperativity in protein folding: from lattice models with side chains to real proteins.** *Fold Des* 1998, **3**:127-139.

65. Pitard E, Garel T, Orland H: **Protein folding, anisotropic collapse, and blue phase.** *J Physique I* 1997, **7**:1201-1210.

66. Barrick D, Baldwin RL: **Three-state analysis of sperm whale apomyoglobin folding.** *Biochemistry* 1993, **32**:3790-3796.

67. Pande VS, Rokhsar DS: **Is the molten globule a third phase of proteins?** *Proc Natl Acad Sci USA* 1998, **95**:1490-1494.

68. Guo Z, Thirumalai D: **Kinetics and thermodynamics of folding of a *de novo* designed four-helix bundle protein.** *J Mol Biol* 1996, **263**:323-343.

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

70. Klimov DK, Thirumalai D: **A criterion that determines the foldability of proteins.** *Phys Rev Lett* 1996, **76**:4070-4073.

71. Camacho CJ, Thirumalai D: **A criterion that determines fast folding of proteins: a model study.** *Europhys Lett* 1996, **35**:627-630.

72. Fukugita M, Lancaster D, Mitchard M: **A heteropolymer model study for the mechanism of protein folding.** *Biopolymers* 1997, **41**:239-250.

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

74. • Klimov DK, Thirumalai D: **Lattice models for proteins reveal multiple folding nuclei for nucleation-collapse mechanism.** *J Mol Biol* 1998, **282**:471-492.
Extensive investigations of the nucleation mechanism using lattice models (of differing lengths and interaction potentials) give support to the multiple folding nuclei model. According to this view, native contacts participate with varying probabilities in folding nuclei. Some of the contacts occur with a higher probability in the transition state than others. The authors show that the average size of the folding nucleus is small, consisting of 10–15 residues.

75. Oxtoby DW: **Nucleation of first-order phase transitions.** *Accounts Chem Res* 1998, **31**:91-97.

76. Arnold U, Ulbrich-Hofmann R: **Kinetic and thermodynamic thermal stabilities of ribonuclease A and ribonuclease B.** *Biochemistry* 1997, **36**:2166-2172.

77. Perl D, Welker C, Schindler T, Schroder K, Marahiel MA, Jaenicke R, Schmid FX: **Conservation of rapid two-state folding in mesophilic, thermophilic, and hyperthermophilic cold shock proteins.** *Nat Struct Biol* 1998, **5**:229-235.

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

79. Thirumalai D, Klimov DK: **Fishing for folding nuclei in lattice models and proteins.** *Fold Des* 1998, **3**:R112-R118.

80. Matheson RR, Scheraga HA: **A method for predicting nucleation sites for protein folding based on hydrophobic contacts.** *Macromolecules* 1978, **11**:814-829.

81. Wolynes PG: **Folding funnels and energy landscapes of larger proteins within the capillarity approximation.** *Proc Natl Acad Sci USA* 1997, **94**:6170-6175.

82. Camacho C, Thirumalai D: **Minimum energy compact structures in random sequences of heteropolymers.** *Phys Rev Lett* 1993, **71**:2505-2508.

83. Thirumalai D, Guo Z: **Nucleation mechanism for protein folding and theoretical predictions for hydrogen-exchange labeling experiments.** *Biopolymers* 1995, **35**:137-140.

84. •• Mantagne A, Radford SE, Dobson CM: **Fast and slow tracks in lysozyme folding: insight into the role of domains in the folding process.** *J Mol Biol* 1997, **267**:1068-1074.
A detailed mapping of the pathways of folding of the hen-egg lysozyme provides direct validation of the kinetic partitioning mechanism. The authors show that 25% ($\Phi = 0.25$) of the molecules fold fast and directly to the native state, whereas the remaining fraction fold by slower processes.

85. Pan J, Thirumalai D, Woodson S: **Folding of RNA involves parallel pathways.** *J Mol Biol* 1997, **273**:7-13.

86. Lazaridis T, Karplus M: **"New view" of protein folding reconciled with the old through multiple unfolding simulations.** *Science* 1997, **252**:672-708.

87. Huenberger PH, Mark AE, Gunsteren WF: **Computational approaches to study protein unfolding: hen egg white lysozyme as a case study.** *Proteins* 1995, **21**:196-213.

88. Duan Y, Wang L, Kollman PA: **The early stage of folding of villin headpiece subdomain observed in a 200-nanosecond fully solvated molecular dynamics simulation.** *Proc Natl Acad Sci USA* 1998, **95**:9897-9902.

89. Dyson HJ, Wright PE: **Peptide conformation and protein folding.** *Curr Opin Struct Biol* 1993, **3**:60-65.

90. Daura X, Jaun B, Seebach D, van Gunsteren WF, Mark AE: **Reversible peptide folding in solution by molecular dynamics simulation.** *J Mol Biol* 1998, **280**:925-932.

91. Matouschek A, Matthews JM, Johnson CM, Fersht AR: **Extrapolation to water of kinetic and equilibrium data for the unfolding of barnase in urea solutions.** *Protein Eng* 1994, **7**:1089-1095.