Pathways for protein folding: is a new view needed?

Vijay S Pande*§, Alexander Yu Grosberg[†][#], Toyoichi Tanaka[†]** and Daniel S Rokhsar*‡

Theoretical studies using simplified models of proteins have shed light on the general heteropolymeric aspects of the folding problem. Recent work has emphasized the statistical aspects of folding pathways. In particular, progress has been made in characterizing the ensemble of transition state conformations and elucidating the role of intermediates. These advances suggest a reconciliation between the new ensemble approaches and the classical view of a folding pathway.

Addresses
*Department of Physics, University of California at Berkeley, Berkeley, CA 94720-7300, USA
†Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
‡Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA; e-mail: rokhsar@physics.berkeley.edu
§e-mail: vijay@physics.berkeley.edu
#e-mail: shura@gels.mit.edu
**e-mail: toyo@tanaka.mit.edu

Abbreviations

C	conformation
CI2	chymotrypsin inhibitor 2
F_{int}	internal free energy
K	total number of contacts
MG	molten globule
N	native state
Q	number of native contacts
P_{fold}	folding probability
R_g	radius of gyration
TS	transition state
U	unfolded state

Introduction

How do proteins fold? While the 35 years since Anfinsen's work have demonstrated the complexity of protein folding, the search continues for the general principles by which proteins adopt their native folds. If such general principles exist, then one might expect them to transcend the specifics of polypeptides. From this point of view, protein folding could be considered a particularly interesting and important case of a more general polymeric phenomenon and therefore much could be learned about the generic aspects of protein folding mechanisms by studying the spontaneous folding of similar polymers. Such relatives of proteins include theoretical cousins that exist only *in silico* or in simplified analytical models.

Here we review recent insights into the kinetics of protein folding that were derived from simplified models of the folding process, considering lattice models for designed heteropolymers (defined below) [1–7,8**,9–16, 17*,18,19], simplified models for real proteins [20–24,25*,26,27,28*] and all-atom molecular dynamics studies [29–34]. These approaches shed light on the nature of folding pathways, their transition states, and the role of intermediates in folding.

We focus on several related issues: the nature of the models and the analysis methods employed in simulations, the mechanism by which chains fold in these simulations, the relationship between kinetics and equilibrium properties, and the importance of conformational entropy in discussing ensembles of conformations. We conclude with a synthesis of the new ensemble-based approaches with the classical pathway picture.

Designed heteropolymers

Before examining the kinetic aspects of simple models of proteins, one must first ask: in what sense are these model heteropolymers protein-like? It is not enough for a polymer to have a unique folded conformation. A heteropolymer with a random sequence will have some lowest energy conformation, and under appropriate temperature and solvent conditions, the polymer will eventually fold to this 'native' state [19,35–40]. But this freezing transition differs from the folding transition in proteins: the folding of random sequences is only weakly cooperative [19,36,37], and proceeds very slowly due to trapping in metastable conformations that are unrelated to the lowest energy conformation [36,40–43]. Also, unlike proteins, the lowest energy conformation of a random sequence of amino acids is expected to be very sensitive to mutations [44–47].

Can sequences be designed to fold in a more protein-like manner? The central goal of all design procedures — both in simple models and with real polypeptides — is to produce sequences with the desired properties, such as fast folding to a stable, preselected native conformation. In this sense, design makes heteropolymers protein-like. (Of course, proteins have other characteristics besides folding — such as specific secondary structures — that cannot be treated adequately in a simplified lattice model.) A general strategy for design is to begin with a collection of random sequences and either select those with the desired property or iteratively improve them. These design strategies have been successfully implemented both theoretically [5,7,10,11,13–16,48] and experimentally [49–54].

One design strategy is to start with a collection of random sequences and select only those that fold in a protein-like

manner. One can then identify the characteristics of these foldable sequences. This approach has been successful both theoretically [55] and experimentally [51]. In [55], the folding of random sequences was simulated *in silico*. It was found that 15% of lattice 27-mers folded reproducibly to their native conformation; these sequences exhibit an energy gap between this native state and other unrelated conformations. In an independent but analogous *in vitro* study [51], a group of random sequence polypeptides were synthesized and proteases were then used to eliminate unfolded sequences; approximately 1% of the sequences remained. Foldable sequences evidently comprise a small but non-negligible fraction of all possible sequences.

A more direct approach to design seeks sequences that fold to a preselected native conformation. To avoid the problems associated with unrelated low energy conformations that act as traps for random sequences, early design schemes looked for sequences with a relatively low internal free energy in the desired native conformation. (For a given conformation (C), the internal free energy, $F_{int}(C)$—often simply referred to as the energy in statistical mechanical models —includes the enthalpy of the polymer as well as the free energy of the polymer–solvent interactions for that specified conformation. In particular, this includes the solvent entropy in the presence of the given conformation and thus incorporates the hydrophobic effect, for further discussion, see [17•].) This approach has proven successful for both lattice models (reviewed in [12,19]) and real proteins (reviewed in [54]). For simple lattice models, it was found that selecting sequences with low native state energy is sufficient to create an energy gap [56]. An important theoretical achievement was the justification of this approach using analytical [46,57] and computational [5,7,43] techniques. Without this understanding, it is unclear why stabilizing a desired fold (and largely ignoring the energy of other conformations) is a sufficient criterion for design.

Phases and free energy landscapes

What are the thermodynamic states of designed heteropolymers? These 'phases'—the unfolded (or denatured) state (U), native state (N), molten globule state (MG) [58,59], and so on—correspond to ensembles of conformations that rapidly interconvert on a time scale (picoseconds) much faster than the typical time scale for protein folding (milliseconds or longer) [60]. The number and nature of these conformations varies for each state; for example, the unfolded state, U, is associated with an enormous number of largely unrelated, unfolded conformations, whereas the native state, N, is associated with a few, closely related, low energy conformations.

Order parameters

A useful way of displaying and conceptualizing the phases of a system is to study the free energy as a function of one or more 'order parameters' — suitably chosen macroscopic quantities that distinguish the different phases. For example, it is common in recent theoretical work to use the number of (tertiary) native contacts in a given conformation, Q, as a macroscopic measure of its folding status. (Two residues are in contact if they are close in space; a common definition requires that the residues' α carbons are within 7 Å of each other). Evidently, Q is a good order parameter in the sense that it distinguishes the unfolded and folded states: unfolded states typically have a small Q, while by definition $Q = Q_{max}$ in the native state.
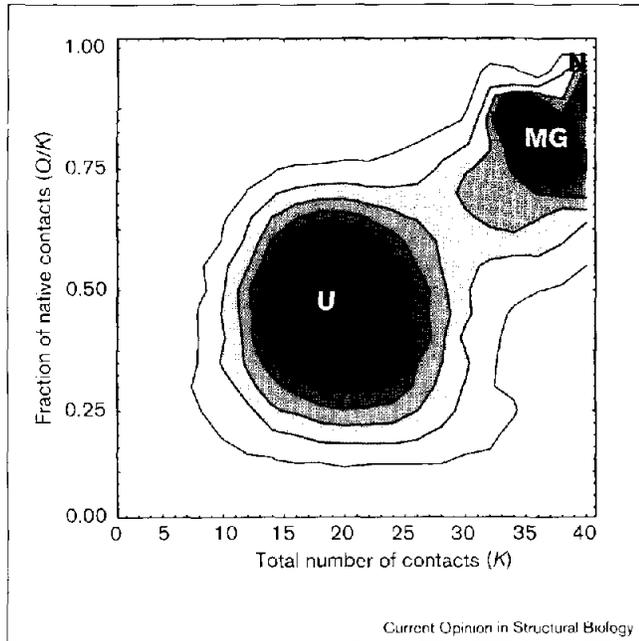
Free energy landscapes

For a simple model polymer, it is straightforward to compute the total free energy as a function of the order parameters. For example, $F_{tot}(Q) = F_{int}(Q) - TS_{conf}(Q)$, where $F_{int}(Q)$ is the average internal free energy of conformations with Q native contacts, and $S_{conf}(Q)$ is the corresponding conformational entropy (roughly the logarithm of the number of accessible conformations with Q native contacts). $S_{conf}(Q)$ is easily computed using Monte Carlo simulations, and is close to zero for the native state and large for the unfolded state.

A plot of the free energy versus one or more order parameters (Figure 1) can be used to describe many aspects of the thermodynamics of the polymer in a quantitative manner:

1. Phases can generally be associated with local free energy minima. This statement makes two hidden assumptions: that the order parameter(s) used are sufficient to distinguish the various phases of the system, and within a minimum, conformations can interconvert rapidly. The value of the order parameter(s) at the minimum describes the nature of the phase (Figure 1).

2. Since different phases respond differently to changes in external conditions, the local minima will shift with temperature, solvent quality, pH and so on. If the interconversion time of conformations within a minimum is fast compared with the transition rate to other minima, we may view this relatively high free energy minimum as a metastable phase. (Such metastable phases are familiar in the case of a supercooled liquid or gas.) One can sometimes make a particular local minimum globally stable by an appropriate choice of external conditions. A good example of this is the stabilization of the MG phase [28•,59] (which is usually either metastable or not present at all under physiological conditions) by the addition of denaturants or a change in pH [61,62].

3. The free energy barrier between two minima indicates distinct phases that are related by a first-order (cooperative) phase transition: when the two minima exchange relative stabilities, the equilibrium value(s) of the order parameter(s) change discontinuously. In contrast, continuous transitions are illustrated by either smooth shifts in the location of a single minimum with changing

external conditions, or the splitting of one minimum into two.

Figure 1

Free energy landscape contours of a 36-mer lattice calculated from a Monte Carlo simulation [28•]. The macroscopic order parameters are the total number of contacts K (both native and non-native) and the fraction of those contacts that are native, Q/K. The three minima correspond to the unfolded (U), molten globule (MG), and native (N) states, respectively. The intervening barriers imply first-order (cooperative) transitions between them. The depths and locations of the minima shift with temperature. Note that the typical unfolded conformation has a substantial number of native contacts; the specific contacts found differ from conformation to conformation [28•].

Funnels

It is important to emphasize the distinction between total free energy surfaces and the heuristic funnel pictures pioneered by Onuchic, Wolynes, Thirumalai et al. [8••,43,63–68,69•], and Chan and Dill [17•,70,71••]. Funnel diagrams plot the internal free energy, F_{int}, (rather than the total free energy) versus unspecified conformation coordinates, and thus do a good job of depicting the energetic (really F_{int}) drive to the native state. This driving force for folding has also been expressed in less picturesque ways [1,3,5,19,42]. In funnel diagrams, conformational entropy is suggested by the width of the funnel in the conformation coordinate. In contrast with the total free energy surfaces discussed above, equilibrium aspects, such as the number of phases, the cooperativity of the transition to the native state, and the relative stability of the phases, are obscured by the funnel visualization, which does not display entropic barriers (only barriers in F_{int}). Finally, applying funnel inspired ideas to kinetics requires knowledge of a good reaction coordinate for folding, which is a difficult and unresolved problem.

Two analogies for folding kinetics

Discussions of protein folding kinetics commonly draw on intuition and terminology from the well-understood theories of chemical reaction rates [72] and the kinetics of first-order phase transitions [73]. Both analogies suggest useful perspectives on the folding problem.

Protein folding as a chemical reaction

Protein folding is often likened to a unimolecular chemical reaction, in which the 'reactant' (an unfolded protein) is converted to a 'product' (the folded state) [74–76]. Unimolecular chemical reactions are typically governed by a single rate-limiting step, when the system passes through the 'transition state'. Intermediate species may or may not be present. Unlike a simple chemical reaction, however, the folding of a polymer is dominated by entropy, in the sense that there are many conformations that correspond to the same stage of the reaction. This has important consequences for the nature of the protein folding pathway, which must therefore be thought of as a sequence of transitions between phases (the unfolded, native, and any intermediate states) rather than individual microscopic conformations.

The transition state of a simple chemical reaction is typically a unique conformation with unfavorable F_{int} that represents the principal barrier between reactants and products [72]. For protein folding, however, the transition state must be regarded as an ensemble of conformations [55], and can only be characterized statistically. Unlike a simple chemical reaction, in which the free energy barrier represents the contribution of a unique conformation, the barrier for protein folding is a total free energy barrier, and may be dominated by conformational entropy (see [18]). Such an entropically generated barrier can be thought of as arising as a result of the relative scarcity of transition state conformations compared with the unfolded state. Thus, the transition state ensemble has many conformations compared with the native state, but many fewer than the unfolded state.

As proteins have many degrees of freedom, in theory there are many different coordinates that could be used to describe the progress of a folding event. Chemical reaction rate theory singles out a particular class of reaction (or transition) coordinates, with the special property of being a slowly varying (preferably the slowest) degree of freedom [72]. The transition state then corresponds to a total free energy maximum along the reaction coordinate. Early theoretical work used Q as a first guess at a reaction coordinate for folding [8••,18,55,69•,77,78]; we will see below that this approach, while qualitatively useful, is fundamentally flawed as a tool for identifying the transition state ensemble [17•].

Protein folding as a first-order phase transition

An even closer analogy may be drawn between the folding of a polypeptide chain into its unique native conformation and the transformation of a vapor into a liquid. In both

cases, there is a dramatic decrease in the conformational entropy of the system that occurs spontaneously upon entering thermodynamic conditions at which the native (or ordered) state has lower free energy. What is the sequence of events — the 'pathway' — by which the ordered free energy minimum is reached?

First order, highly cooperative, bulk phase transitions typically proceed via a nucleation and growth mechanism [73]. Consider a disordered phase (for example, gas) that is suddenly quenched to a low temperature at which a more ordered phase (for example, liquid) has lower free energy. At this low temperature, the disordered phase is only metastable (or 'supercooled'). Thermal fluctuations in the microscopic conformation of the system lead to the spontaneous formation and dissolution of small droplets

of the more ordered phase floating within the metastable disordered state. Once a critical nucleus — a droplet of critical radius $R^{\ddagger}$ — forms (Figure 2), however, it can grow rapidly by accretion, driven by the overall thermodynamic stability of the condensed state. Thus the critical nucleus can be thought of as the transition state for the vapor to liquid transition.

An essential feature of this scenario is that only local free energy barriers need to be surmounted. That is, the crucial event is the formation of a small ordered droplet, which is a local process. Spontaneous thermal fluctuations — a search among microstates — need only find a critical droplet, not the completely ordered state. For proteins, this would correspond to a search for one of the many members of the transition state ensemble rather than a Levinthal-like

Figure 2
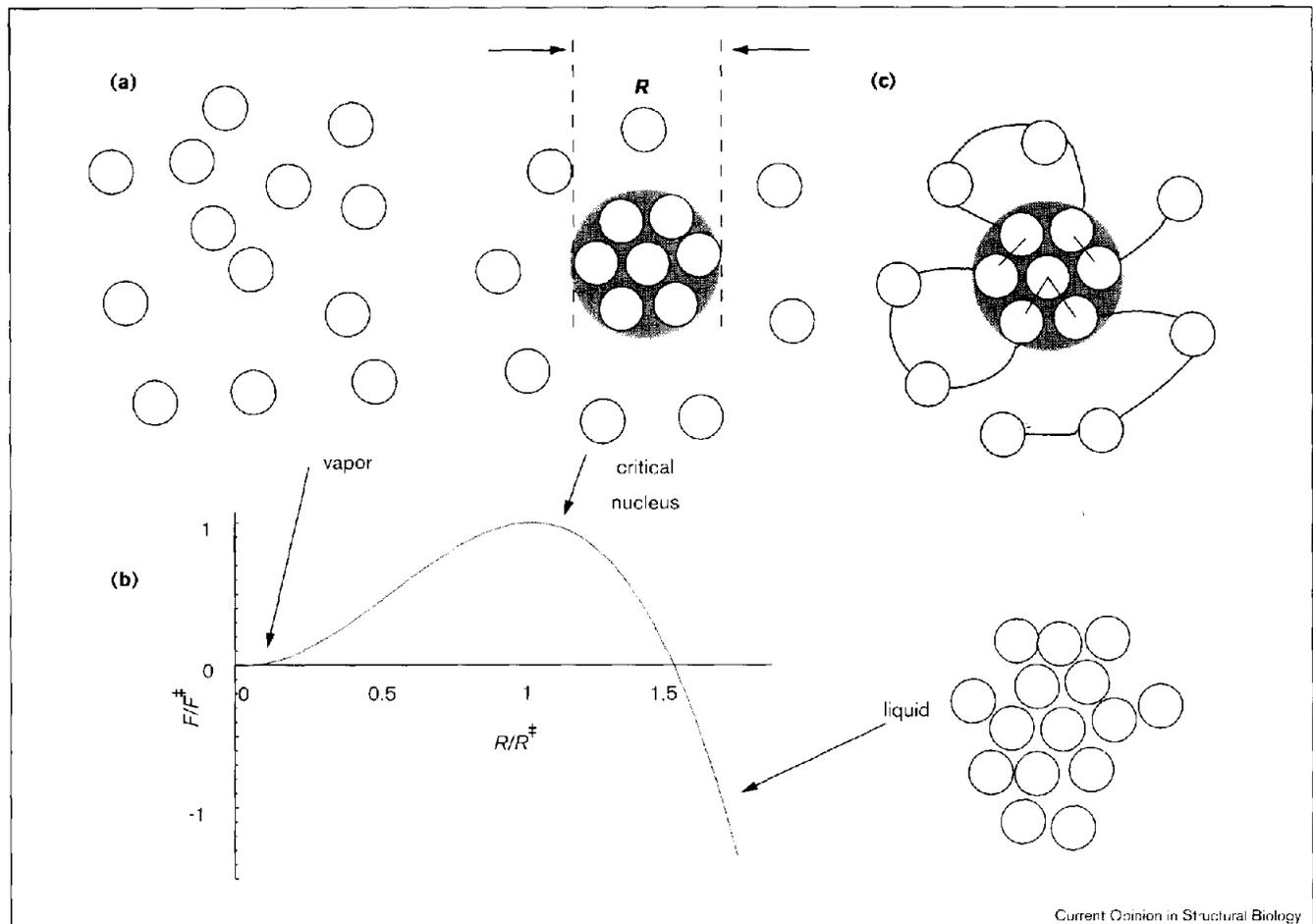
Nucleation mechanisms in bulk transitions and polymers. **(a)** First-order (cooperative) phase transitions proceed by a nucleation mechanism in which a small droplet of the ordered phase is formed within the metastable disordered phase. **(b)** The free energy $F(R)$ of a droplet depends on its radius R. There is a free energy gain proportional to the volume of the droplet, $-\delta f R^{3}$, where δf is the free energy difference per unit volume between the ordered and disordered phases. Opposing this free energy gain is the cost $-\gamma R^{2}$ of the surface of the droplet, which is the product of the surface tension γ (the interfacial free energy per unit area) between the two phases, and the surface area of the droplet. The net free energy of the condensed droplet then has a free energy maximum or barrier near $R^{\ddagger} \sim \delta f/\gamma$ where the bulk gain begins to offset the surface cost. **(c)** A similar mechanism could apply to the folding of a protein, with the ordered phase identified with the native state and the disordered phase identified with the unfolded phase. This droplet would represent the transition state. Different conformations of the loops correspond to different members of the transition state ensemble.

search for the unique native conformation. This simple homogeneous nucleation picture of a bulk transition will of course need to be modified to account for the polymeric chain connectivity (T Garel, H Orland, E Pitard, personal communciation; [79–82]) and the heterogeneity of the sequence, which may favor specific droplets over others.

Implications for protein folding

These two analogies shape our understanding of the process of protein folding. What do they teach us? If we view folding as a two-state transition — a chemical reaction characterized by a single kinetic phase — then we expect to find a well-defined transition state ensemble. This ensemble could be easily characterized if we knew the appropriate reaction coordinate for folding. From the theory of first-order phase transitions, however, we expect the rate-limiting step of protein folding to be the formation of some structure or structures analogous to the critical nucleus, which plays the role of transition state for a first-order transition.

The physical picture of a first-order transition demonstrates that an order parameter for an equilibrium transition (for example, Q) is not necessarily useful for determining the transition state ensemble that controls kinetics. In particular, the order parameter Q is a poor reaction coordinate for folding because it measures a global property — the total number of native contacts — and is therefore not sensitive to the distribution of those contacts. (The same statement applies to the radius of gyration parameter, R_g.) Yet from the study of the liquid to gas transition we see that the spatial distribution of the ordered phase (droplets) within the disordered state is a central aspect of the mechanism. We therefore expect that the net amount of order (the number of native contacts or the total volume of the condensed phase) will not be a good reaction coordinate. Rather, the search for a proper reaction coordinate for folding must acknowledge the possibilty that the transition state contains some sort of local structure. But how do we identify a reaction coordinate without already knowing the nature of the transition state?

Du *et al.* [83••] have recently proposed a straightforward (but computationally intensive) procedure for determining transition states without making any assumptions about the reaction coordinate. Their approach therefore allows an unbiased analysis of the transition state ensemble (VS Pande, DS Rokhsar, unpublished data). They introduce the folding probability, $p_{fold}(C)$, which measures the probability that a simulation starting from conformation C will reach the folded state before encountering an unfolded conformation. If C is very close to the native conformation, then $p_{fold} \approx 1$; if C is near the unfolded phase, then $p_{fold} \approx 0$.

The transition state ensemble consists of those conformations sampled during a folding event whose $p_{fold} = 1/2$,

that is conformations that are equally likely to fold or unfold. (For a two-state transition, there is a single, well-defined transition state. If there are intermediates, then $p_{fold} \approx 1/2$ determines the major transition state that governs the rate-limiting step.) The relative weight of a conformation in the transition state ensemble is defined by its rate of appearance in folding events. In general, conformations with $p_{fold} = 1/2$ do not appear with equal weight in the transition state ensemble (VS Pande, DS Rokhsar, unpublished data). The p_{fold} method allows individual transition state conformations to be unambiguously identified for a given folding trajectory without making any assumptions regarding the reaction coordinate, and is particularly useful in the absence of a valid reaction coordinate.

Nature of the transition state ensemble

It is widely believed that some sort of nucleation event is central to the mechanism of protein folding (VI Abkevich, LA Mirny, EI Shakhnovich, personal communication; [23,38,84••, 85,86••,87–95]), although the detailed nature of this nucleation mechanism is still under debate. In this section, we review recent simulations of designed lattice heteropolymers that address the transition state for protein folding. In this discussion we emphasize the connection between the simulation methodology employed (that is, how the transition state ensemble is determined) and the resulting picture illustrating how proteins fold. There are three competing scenarios.

Evidence for many delocalized nuclei

One scenario envisions the transition state ensemble as consisting of many delocalized nuclei [8••]. That is, each conformation in the transition state ensemble contains a different locally structured region or nucleus reminiscent of the jigsaw model postulated by Harrison and Durbin [96]. This theory is supported by the work of Onuchic, Socci, Luthey-Schulten and Wolynes [8••,69•], who investigated the transition state ensemble in 27-mers. They used Q as a reaction coordinate, computed the total free energy $F_{tot}(Q)$, and identified the $Q_{barrier}$ at which $F_{tot}(Q)$ has a maximum. If Q were a good reaction coordinate, conformations with $Q = Q_{barrier}$ would comprise the transition state ensemble. The fact that this analysis is based on the problematic assumption that Q is a valid reaction coordinate does not necessarily rule out the resulting physical picture.

Onuchic *et al.* [8••] conclude that the transition state ensemble comprises many partially folded conformations. A 27-mer has a total of 10^{16} possible conformations, with 10^{10} of them semicompact or highly collapsed [55]. The authors estimate that 10^4 of these conformations make up the transition state ensemble, with $Q \approx 0.6\ Q_{max}$. That is, they suggest that a typical transition state conformation contains 60% of the contacts found in the native state. Furthermore, "different native contacts have different

degrees of participation" in the transition state, hence the use of the term delocalized.

In an earlier study, Sali, Shakhnovich and Karplus [55] also examined the folding of designed 27-mers. Also using Q as a reaction coordinate, they identified a different barrier in $F_{tot}(Q)$ as the major transition state, and inferred that the transition state ensemble consists of all 10^3 semicompact conformations, with $0.8 \leq Q/Q_{max} \leq 1$. The high fraction of native contacts implies that the transition state is very close to the folded lattice conformation. Since these conformations are all different (except for their common resemblance to the native state), these results have been cited [97] as evidence of many parallel folding pathways.

Recent work by Chan and Dill [17•] has also emphasized the possibility that the transition state ensemble involves a diverse collection of largely unrelated conformations: "since the idea of transition state is really about rate limits and bottlenecks, it includes all the conformations that are passed through on the way to the native state, because they are all responsible for determining the rate" [71••]. Recognizing that Q is not a suitable reaction coordinate for folding, they introduce a novel kinetic reaction coordinate for lattice models that corresponds to the minimum number of steps needed to reach the native state from a given conformation, following a minimum energy path. They find that the transition state is not characterized by a specific bottleneck structure, but rather a broad ensemble lacking specific structure.

Evidence for a specific nucleus

Based on their analysis of the folding of a designed 36-mer, a qualitatively different kind of transition state ensemble was proposed by Abkevich, Gutin and Shakhnovich [85]. They found that specific core native contacts were reproducibly formed early in folding. Moreover, once these particular contacts are formed, folding proceeds rapidly. Their results suggest that the transition state ensemble comprises conformations that share the same set of essential contacts which then form a compact core inside the native state — a specific nucleus (VI Abkevich, LA Mirny, EI Shakhnovich, personal communication; [85,86••]). As a test of this hypothesis, Shakhnovich *et al.* [86••] have confirmed that different sequences designed for the chymotrypsin inhibitor 2 (CI2) backbone have conserved residues at the predicted core positions. This picture closely resembles nucleation in first-order phase transitions, with a critical nucleus specified by the heterogeneity of the polymer.

If the presence of these specific contacts is the only requirement for a conformation to be found in the transition state ensemble, then this ensemble would comprise related conformations that differ only in the configuration of the polymeric loops that lie between core contacts (Figure 2c). Thus, despite the formation of an ordered core, the transition state ensemble in the

specific nucleus has a substantial entropy arising from the conformational freedom of these loops.

Evidence for transition state classes

A third scenario is proposed by Pande and Rokhsar (unpublished data), who analyzed folding pathways and the transition state ensembles for a range of polymer lengths from 27- to 64-mers. They directly determined the transition state ensemble using the p_{fold} method, thereby avoiding ambiguities associated with the choice of reaction coordinate. The transition state ensemble is defined by collecting p_{fold} $(C) = 1/2$ conformations from several hundred folding trajectories, using the same sequence, but starting from different unfolded conformations. For 27-mers, the transition state ensemble consists of a collection of closely related conformations — a single class — that share a specific set of core contacts with high probability, and other selected optional contacts with intermediate probability. For longer chains, the transition state ensemble may consist of a few distinct classes.

As for the specific nucleus picture, the conformational freedom of the loops endows the transition state with a large entropy. Pande and Rokhsar emphasize that the entropy of a transition state class is further enhanced by the combinatorial possibilities for choosing the optional contacts. Indeed, this value is large (typically 10^9 conformations for a 48-mer) and therefore cannot be ignored. For longer polymers, the transition state ensemble of a typical designed heteropolymer contains two or three such classes but the transition state ensemble of fast-folding sequences (VI Abkevich, LA Mirny, EI Shakhnovich, personal communication) consists of a single class (VS Pande, DS Rokhsar, unpublished data).

Which physical picture is correct?

We have seen that recent theoretical work suggests three distinct physical pictures of the transition state ensemble: many delocalized nuclei, a specific nucleus, and transition state classes. Which of these possibilities applies to protein folding? While the most recent simulations (VS Pande, DS Rokhsar, unpublished data) using the p_{fold} method [83••] support the transition state class scenario in lattice models, the nature of the transition state ensemble in real proteins is best addressed experimentally.

Φ-value analysis

The principal experimental method for identifying transition states for folding is the Φ analysis introduced by Fersht [84••,91]. Site-directed mutagenesis was used to perturb both the transition and native states. Then $\Phi \equiv \Delta(G_{\ddagger} - G_U)/\Delta(G_N - G_U)$ measures the degree to which the free energy of the transition state is affected relative to the native state. (The Δ term refers to the difference between the mutant and wild-type proteins.) A residue that participates in the same interactions in both the native and transition states would ideally have $\Phi = 1$, whereas a residue with $\Phi = 0$ is likely to be unstructured

in the transition state. In practice, one also finds fractional Φ-values, which can be interpreted in two ways: either the residue makes native-like contacts in only a fraction of the transition state conformations, or the residue makes contacts in the transition state ensemble that are weakened relative to those it forms in the native state. Fersht and colleagues [97,98] favor the second interpretation based on a comparison of single versus multiple pathway models with kinetic data.

Comparing theory with experiment

How do the pictures derived from simple theoretical models compare with these experimental results? The principal focus is the explanation of fractional Φ values. A histogram [8••] of experimentally determined Φ values for CI2 [91,99] is broadly peaked between Φ = 0 and 0.6. Onuchic *et al.* [8••] find a similarly broad distribution of Φ-value analogs for a 27-mer lattice (which they argue is comparable to a 60-residue protein). The molecular dynamics sampling of fragment B from protein A by Boczko and Brooks [33] yields a qualitatively similar distribution (reported in [8••]). Onuchic *et al.* use these broad distributions to support their many delocalized nuclei picture. They note that if the strict specific nucleus picture were valid, the Φ-value probability distribution would be bimodal — residues in the nucleus would have a high Φ value, while residues not in the nucleus should have Φ ~ 0.

The simulations of Pande and Rokhsar (unpublished data) also give a broad distribution of Φ-value analogs (the fraction of transition state conformations that possess a given native contact). Unlike Onuchic *et al.* [8••], however, they explain the broad distribution of Φ values by addressing the variation between conformations within a transition state class. Their required core contacts have high Φ values, while the optional contacts have lower values.

Which interpretation is correct? Fersht *et al.* [97] rule out multiple pathways for CI2 by referring to a Bronsted analysis in which the logarithm of the folding (or unfolding) rate is plotted against the destabilization of the folded state for a series of mutants. These plots are linear, suggesting that the reaction kinetics can be modeled by a single class of transition state. A similar analysis for the larger protein barnase suggests, however, that this may not be a general result [98].

Molecular dynamics simulations of unfolding at high temperature

All-atom simulations of unfolding trajectories of CI2 under extreme conditions (500K, 26 atmospheres) conducted by Daggett *et al.* [30,31] may also shed light on the nature of the transition state ensemble. Under these conditions, unfolding is accelerated by six orders of magnitude, from milliseconds to nanoseconds, and becomes accessible

Figure 3
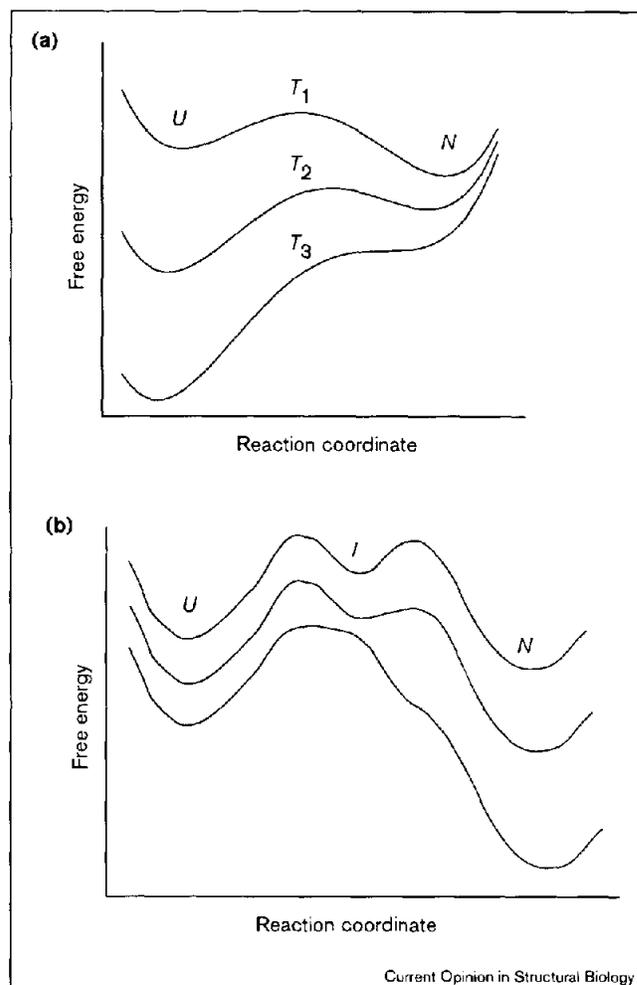
Temperature dependence of barriers and intermediates. **(a)** This schematic plot of total free energy versus a generic reaction coordinate for several temperatures ($T_3 \gg T_2 > T_1$) illustrates that barriers are temperature dependent. Under extreme conditions ($T_3 \gg T_2$), the free energy barrier may disappear. **(b)** Similarly, the presence or absence of a metastable intermediate (*I*) may depend on temperature. Although these pictures are schematics, they are based on real simulation data (VS Pande, DS Rokhsar, unpublished data).

to study. They argue that the transition state should correspond to a rapid change in the conformation of the protein with time, and identify related conformations in four unfolding trajectories as putative transition states. One might argue that the transition state for unfolding under extreme conditions could be quite different from the transition state under more standard conditions. In particular, an entropically generated free energy barrier of the sort found in lattice models may not even be present under extreme temperature and pressure conditions if the native state loses its metastability (Figure 3a). Nevertheless, there is remarkable consistency between the residues Daggett *et al.* [30,31] identify as important in the transition state and those implicated experimentally by Fersht and co-workers [99] using Φ-value analysis.

Designing pathways

A complete understanding of the mechanism(s) of protein folding should include a procedure for redesigning folding pathways. That is, in addition to designing the equilibrium properties of a heteropolymer, one should be able to intentionally manipulate its folding kinetics. As a first step in this direction, Abkevich, Mirny, and Shakhnovich (personal communication) have used an evolution-like process to select fast folding lattice heteropolymer sequences by mutating sequences and retaining those variants that fold most quickly.

Abkevich, Mirny and Shakhnovich (VI Abkevich, LA Mirny, EI Shakhnovich, personal communication) find that all fast folding sequences designed in this manner fold with the same specific nucleus. An analysis by Pande and Rokhsar (unpublished data) of the pathways of these fast-folding sequences using the p_{fold} method shows that they fold via a single transition state class that is energetically preferred among the several possible classes of typical heteropolymers designed for equilibrium folding to a specific native state conformation. That is, evolutionary design for fast folding leads to a specific pathway. Pande and Rokhsar (unpublished data) have used this idea to directly design sequences (that is, without an evolutionary selection for fast folding) with both a preselected native state conformation and a chosen transition state class. The fact that the transition states can be manipulated in this manner supports the specific nucleus and transition state class pictures, but is more difficult to reconcile with the many delocalized nuclei scenario.

Intermediates

Many small proteins fold without detectable intermediates [88,100–102]. Yet there are clear examples of other proteins whose folding route passes through partially folded, MG-like, on-pathway intermediates [61,103–106]. Furthermore, other proteins fold with so-called off-pathway intermediates that are in some sense misfolded, most notably those involving proline isomerization [74] and/or disulfide bond rearrangements [107]. Such intermediates are either inferred from multistate kinetics or trapped using a variety of experimental techniques.

Some recent lattice and off-lattice studies have found both on- and off-pathway intermediates in direct simulations of folding events. Other studies have not revealed such intermediates, which may be due either to differences in the methodologies of the different calculations (for example, different temperatures of the simulations) or to real variations between the folding pathways of different sequences. Perhaps the only general statement that can be made is that if intermediates are metastable phases of the polymer (that is, locally stable minima of the free energy surface; Figure 1), then as the folding temperature, pressure, and pH are varied the stability of such a state will change and it may disappear (Figure 3b). Thus, the presence or absence of intermediates for any given protein is likely to be sensitive to folding conditions.

In their lattice simulations, Pande and Rokhsar (unpublished data) found that each on-pathway, partially folded intermediate is associated with a corresponding transition state class. The conformations that comprise the intermediate state contain a common frozen core of contacts, surrounded by fluctuating loops. The conformational entropy of the loops stabilizes the intermediate, which is a metastable phase. Pande and Rokhsar demonstrate this directly by computing the free energy surface with respect to two order parameters, the number of native contacts Q and the number of core contacts Q_{core}. $F(Q,Q_{core})$ exhibits a metastable intermediate minimum along with the unfolded and folded minima. At sufficiently low temperatures, the barrier between the intermediate and native state disappears, and the transition becomes two-state (Figure 3).

Mirny, Abkevich, and Shakhnovich [108] discovered that well-designed sequences are more stable in their native state and fold quickly without intermediates in a two-state process. Less-optimized sequences, however, fold more slowly, via parallel pathways involving misfolded intermediates.

Off-pathway intermediates have been found in coarse-grained, nonlattice models of four-helix bundles studied by Thirumalai and co-workers [26,43]. They performed Langevin dynamics simulations in which the polypeptide was modeled by chain of spheres (representing the α carbons) connected by springs, using a three-letter code to indicate hydrophobic, polar, and neutral residues. They found intermediates that are misfolded (one of the helices being kinked) and show that folding is accelerated if the intermediate is destabilized [26]. This work also suggests that intermediates can be regarded as metastable, equilibrium phases.

Boczko and Brooks [33] studied the thermodynamic properties of a small three-helix bundle (fragment B of protein A) using an all-atom approach. They simulated approximately 10 ns of unfolding, at a variety of temperatures (ranging from 300K to 400K), and sampled conformations at many values of R_g in order to piece together the free energy $G(R_g)$. Conformations generated in this run were used to construct clusters with a given R_g. From this analysis, they inferred a folding intermediate for this small protein. Recent experiments on protein A [109], however, may contradict these results.

Folding pathways

The classical view of folding envisions a defined sequence of states leading from the unfolded to the native state, allowing for the possibility of on-pathway (partly folded) or off-pathway (misfolded) intermediates [74,75]. Several years ago, Baldwin [110,111] suggested that a new view

of folding was emerging based on simplified statistical mechanical models for proteins. As we have seen, these models emphasize both ensemble properties and the importance of pathways without intermediates for rapidly folding proteins. More recently, the term 'new view' has acquired a broader meaning [43,71••] that stresses the possibility of a diverse myriad of pathways with delocalized transition states. According to this approach [8••,17•,43,69•,71••], the central feature of this new view is the replacement of the pathway concept with picturesque funnel diagrams that illustrate features of protein folding and the role of ensembles. This funnelist viewpoint has recently been reviewed in detail by Dill and Chan [71••].

In contrast, other studies of simple models (VS Pande, DS Rokhsar, unpublished data; [56]) do not suggest a radical new view, but rather a refinement of the classical picture in which the classical concepts of states and pathways are interpreted in terms of ensembles of conformations. For example, each step in a classical pathway can be precisely regarded as a transition between two phases (ensembles of rapidly interconverting conformations [60]), so that folding proceeds through a sequence of metastable phases. In the next section, we briefly summarize what might be called a 'neo-classical' view, an alternative to the funnel picture.

Classical pathways from an ensemble view

Thirty five years of studying protein folding kinetics has shown that folding reactions can be analyzed using pathways of varying complexity, such as the two-state model

$$U \rightleftharpoons N \tag{1}$$

or models with on- or off-pathway intermediates,

$$U \rightleftharpoons I_{on} \rightleftharpoons N, \quad I_{off} \rightleftharpoons U \rightleftharpoons N \tag{2}$$

and so on. Increasingly complex schemes become increasingly difficult to compare with experimental results, and there are, as yet, no first principles rules to determine in advance which pathway applies to a specific protein.

In simple chemical reactions, the symbols in the mass action equations (1) and (2) represent specific conformations of a small molecule. For protein folding, however, we must interpret each symbol as an ensemble of rapidly interconverting conformations — that is, a thermodynamic phase. In a temperature jump experiment, for example, U would be a supercooled (metastable) phase (since the unfolded state is not thermodynamically stable at the refolding temperature), and N would be the stable native state. Intermediates, if present, appear as metastable phases; as we have seen, some recent simulations exhibit intermediates that appear to be metastable, MG-like phases (VS Pande, DS Rokhsar, unpublished data; [28•]) others exhibit misfolded, off-pathway intermediates [26,108].

In this ensemble view of a classical pathway, the $\rightleftharpoons$ arrows denote first-order, cooperative phase transitions. The free energy barriers between phases are surmounted by passing through a well defined ensemble of transition state conformations. In a two-state reaction, the rate-limiting step is the attainment of the transition state between the initial and final states. Once a member of the transition state ensemble has been reached, folding can occur rapidly.

This extension of the classical pathway idea provides a very different physical picture from the funnelist viewpoint, which replaces the chemical reaction analogy with a picture of the conformations streaming down an internal free energetic funnel that directs each conformation towards the native state [43,71••]. These two scenarios — the new view based on funnels and the neo-classical view based on transitions between phases — are distinct physical pictures of the folding process. Experiments and simulations must ultimately choose between them.

Levinthal revisited

By what process does an unfolded polymer reach a transition state conformation? Levinthal argued that a random search among conformations would never find the native state ([112,113]; discussed in [114]). While this is true, it is also irrelevant: the randomly fluctuating unfolded polypeptide only needs to find one of the many members of the transition state ensemble, not a unique conformation. To test the random search for the transition state ensemble hypothesis, one can compare the folding time to that estimated for a random search [55]: $t_f^{random} = t_0(W_U/W_{TS})$ — the typical time taken to sample a distinct conformation (t_0) multiplied by the ratio of the number of unfolded states (W_U) to the number of transition states (W_{TS}).

Using Q as a reaction coordinate to describe the transition state, Sali, Shakhnovich and Karplus [55,77] suggested that in 27-mer lattice models, the polymer finds a member of the transition state ensemble by a random search. Using the more reliable p_{fold} approach [83••] and longer chains, Pande and Rokhsar (unpublished data) have demonstrated the existence of a random search mechanism using two independent means. First, they found that the conformations sampled in the unfolded state were uncorrelated. Second, they found that the mean first passage folding times measured using Monte Carlo simulations agree with the calculation of t_f^{random}-employing simulation measurements of W_U and W_{TS}. (The combinatorial entropy of the optional transition state contacts is critical for this agreement.)

Conclusions

Recent theoretical developments using simplified models have brought about an increased awareness of the importance of ensembles in understanding the folding process. But have these new models actually led to a new view of folding? The principal advantage of the new models is that the nature of the folding pathways can, in

principle, be completely understood by direct simulation of folding on a computer, where every detail is accessible. We have seen that the conformation by conformation trajectory of the polymer can be understood in terms of ensembles of rapidly interconverting conformations or phases of the polymer. These ensembles can be identified directly in simple models, which permit a complete analysis of the unfolded state, transition state ensemble, and intermediates, as discussed above. Thus, in these new models, the folding pathway can be dissected in microscopic detail.

We have argued that these new models do not require a 'new view' of folding. Protein folding can be understood by extending the classical view to include ensembles in a natural fashion. In this sense, some of the new statistical approaches to the folding process are perhaps better characterized as 'neo-classical' rather than a fundamentally new alternative. Pathways for folding imply "a well-defined sequence of events which follow one another" [112], where an event should be interpreted as a transition from one phase to another. The nature of these transitions has been clarified by the study of simple models that focus on the essential heteropolymeric aspects of the folding process. As this emerging neo-classical view develops, we look for increasing comparisons with experiments, the ultimate arbiter of theoretical progress.

Acknowledgements

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. Ueda Y, Taketomi H, Go N: Studies on protein folding, unfolding and fluctuations by computer simulation. I. The effect of specific amino acid sequence represented by specific inter-unit interactions. *Int J Peptide Res* 1975, 7:445-459.

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

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

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

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

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

7. Pande VS, Grosberg A Yu, Tanaka T: Thermodynamic procedure to synthesize heteropolymers that can renature to recognize a given target molecule. *Proc Natl Acad Sci USA* 1994, 91:12976-12981.

8. 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.
Simulations of 27-mer folding support the many delocalized nuclei scenario of the transition state ensemble, which is estimated by assuming that the number of native contacts (Q) is an appropriate reaction coordinate. The funnel picture is reviewed and refined.

9. Chan HS, Dill KA: Comparing folding codes for proteins and polymers. *Proteins* 1996, 24:335-344.

10. Hinds DA, Levitt M: From structure to sequence and back again. *J Mol Biol* 1996, 258:201-209.

11. Abkevich VI, Gutin AM, Shakhnovich EI: Improved design of stable and fast-folding model proteins. *Fold Des* 1996, 1:221-230.

12. Shakhnovich EI: Modeling protein folding: the beauty and power of simplicity. *Fold Des* 1996, 1:R50-R52.

13. Morrissey MP, Shakhnovich EI: Design of proteins with selected thermal properties. *Fold Des* 1996, 1:391-405.

14. Olszewski KA, Kolinski A, Skolnick J: Folding simulations and computer redesign of protein A three-helix bundle motifs. *Proteins* 1996, 25:286-299.

15. Kurosky T, Deutsch JM: Design of copolymeric materials. *J Phys-A Math Gen* 1996, 27:L387-L389.

16. Deutsch JM, Kurosky T: New algorithm for protein design. *Phys Rev Lett* 1996, 76:323-326.

17. Chan HS, Dill K: Protein folding kinetics from the perspective
• of simple models. *Proteins* 1997, 8:2-33.
This paper developes a novel kinetic reaction coordinate, and describes a scenario in which the transition state for a protein folding is a broad ensemble of conformations lacking specific structure.

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

19. Pande VS, Grosberg A Yu, Tanaka T: Heteropolymer freezing and design: towards physical models of protein folding. *Rev Mod Phys* 1998, in press.

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

21. Skolnick J, Kolinski A: Dynamic Monte Carlo simulations of globular protein folding/unfolding pathways. I. Six-member, Greek key beta-barrel proteins. *J Mol Biol* 1990, 212:787-817.

22. Rey A, Skolnick J: Computer modeling and folding of four-helix bundles. *Proteins* 1993, 16:8-28.

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

24. Kolinski A, Galazka W, Skolnick J: On the origin of the cooperativity of protein folding: implications from model simulations. *Proteins* 1996, 26:271-287.

25. 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.
A simplified representation of a polypeptide is introduced in order to simulate the folding kinetics of a four-helix bundle. Misfolded intermediates are also identified.

26. Camacho CJ, Thirumalai D: Denaturants can accelerate folding rates in a class of globular proteins. *Protein Sci* 1996, 5:1826-1832.

27. Veitshans T, Klimov D, Thirumalai D: Protein folding kinetics: timescales, pathways, and energy landscapes in terms of sequence-dependent properties. *Fold Des* 1997, 2:1-22.

28. Pande VS, Rokhsar DS: Is the molten globule a third phase of
• proteins? *Proc Natl Acad Sci USA* 1998, in press.
The molten globule phase is found in lattice and off-lattice protein models and it is thought that this phase is a general heteropolymer property. The stability of the molten globule is discussed in terms of the conformational entropy of disordered loops.

29. Caflisch A, Karplus M: Molecular dynamics simulation of protein denaturation: solvation of the hydrophobic core and secondary structure of barnase. *Proc Natl Acad Sci USA* 1994, 91:1746-1750.

30.   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.

31.   Alonso DO, Daggett V: **Molecular dynamics simulations of protein unfolding and limited refolding: characterization of partially unfolded states of ubiquitin in 60% methanol and in water.** *J Mol Biol* 1995, 247:501-520.

32.   Van Gunsteren WF, Hunenberger PH, Kovacs H, Mark AE, Schiffer CA. **Investigation of protein unfolding and stability by computer simulation.** *Philos Trans R Soc Lond Biol* 1995, 348:49-59.

33.   Boczko EM, Brooks CL III: **First-principles calculation of the folding free energy of a three-helix bundle protein.** *Science* 1995, 269:393-396.

34.   Tirado-Rives J, Orozco M, Jorgensen WL: **Molecular dynamics simulations of the unfolding of barnase in water and 8 M aqueous urea.** *Biochemistry* 1997, 36:7313-7329.

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

36.   Sfatos CD, Gutin AM, Shakhnovich EI: **Phase diagram of random copolymers.** *Phys Rev E* 1993, 48:465-475.

37.   Pande VS, Grosberg A Yu, Tanaka T: **Freezing transition of random heteropolymers consisting of an arbitrary set of monomers.** *Phys Rev E* 1995, 51:3381-3402.

38.   Gutin AM, Abkevich VI, Shakhnovich EI: **Chain length scaling of protein folding time.** *Phys Rev Lett* 1996, 77:5433-5436.

39.   Wang J, Onuchic JN, Wolynes PG: **Statistics of kinetic pathways on biased rough energy landscapes with applications to protein folding.** *Phys Rev Lett* 1996, 76:4861-4864.

40.   Wang J, Plotkin SS, Wolynes PG: **Configurational diffusion on a locally connected correlated energy landscape: application to finite, random heteropolymers.** *J Phys I* 1997, 7:395-421.

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

42.   Pande VS, Grosberg A Yu, Joerg C, Tanaka T: **How well is heteropolymer freezing described by the random energy model?** *Phys Rev Lett* 1996, 76:3987-3990.

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

44.   Shakhnovich EI, Gutin AM: **Influence of point mutations on protein structure: probability of a neutral mutation.** *J Theor Biol* 1991, 149:537-546.

45.   Bryngelson J: **When is a potential accurate enough for structure prediction? Theory and application to a random heteropolymer model of protein folding.** *J Chem Phys* 1994, 100:6038-6045.

46.   Pande VS, Grosberg A Yu, Tanaka T: **How accurate must potentials be for successful modeling of protein folding?** *J Chem Phys* 1995, 103:9482-9491.

47.   Bussemaker HJ, Thirumali D, Bhattacherjee JK: **Thermodynamic stability of folded proteins against mutations.** *Phys Rev Lett* 1997, 79:3530-3533.

48.   Shakhnovich EI: **Proteins with selected sequences fold into unique native conformation.** *Phys Rev Lett* 1994, 72:3907-3910.

49.   Regan L, DeGrado WR: **Characterization of a helical protein designed from first principles.** *Science* 1988, 241:976-978.

50.   Hecht MH, Richardson JS, Richardson DC, Ogden RC: **De novo design, expression, and characterization of Felix: a four-helix bundle protein if native-like sequence.** *Science* 1990, 249:884-891.

51.   Davidson AR, Lumb KJ, Sauer RT: **Cooperatively folded proteins in random sequence libraries.** *Nat Struct Biol* 1995, 2:856-863.

52.   Dahiyat BI, Mayo SL: **Protein design automation.** *Protein Sci* 1996, 5:895-903.

53.   Lazar GA, Desjarlais JR, Handel TM: **De novo design of the hydrophobic core of ubiquitin.** *Protein Sci* 1997, 6:1167-1178.

54.   DeGrado WF: **Proteins from scratch.** *Science* 1997, 278:80-81.

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

56.   Shakhnovich EI: **Theoretical studies of protein-folding thermodynamics and kinetics.** *Curr Opin Struct Biol* 1997, 7:29-40.

57.   Ramanathan S, Shakhnovich E: **Statistical mechanics of proteins with 'evolutionary selected' sequences.** *Phys Rev E* 1994, 50:1303-1312.

58.   Ptitsyn OB, Uversky VN: **The molten globule is a third thermodynamical state of protein molecules.** *FEBS Lett* 1994, 341:15-18.

59.   Uversky VN, Ptitsyn OB: **All-or-none solvent-induced transitions between native, molten globule and unfolded states in globular proteins.** *Fold Des* 1996, 1:117-122.

60.   Zwanzig R: **Two-state models of protein folding kinetics.** *Proc Natl Acad Sci USA* 1997, 94:148-150.

61.   Ptitsyn OB: **Structures of folding intermediates.** *Curr Opin Struct Biol* 1995, 5:74-78.

62.   Haynie DT, Freire E: **Structural energetics of the molten globule state.** *Proteins* 1993, 16:115-140.

63.   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.

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

65.   Onuchic JN, Wolynes PG, Luthey-Shulten Z, Socci ND: **Towards an outline of the topography of a realistic protein folding funnel.** *Proc Natl Acad Sci USA* 1995, 92:3626-3630.

66.   Shoemaker BA, Wang J, Wolynes PG: **Structural correlations in protein folding funnels.** *Proc Natl Acad Sci USA* 1997 94:777-782.

67.   Plotkin SS, Wang J, Wolynes PG: **Correlated energy landscape model for finite random heteropolymers.** *Phys Rev E* 1996, 53:6271-6296.

68.   Plotkin SS, Wang J, Wolynes PG: **Statistical mechanics of a correlated energy landscape model for protein folding funnels.** *J Chem Phys* 1997, 106:2932-2948.

69.   Socci ND, Onuchic JN, Wolynes PG: **Diffusive dynamics of the**
•   **reaction coordinate for protein folding funnels.** *J Chem Phys* 1996, 104:5860-5868.
See also [43], in which the authors describe protein folding in terms of a folding funnel. The essential role of an internal free energetic drive towards the native state is emphasized and the role of intermediates in the funnel picture is reviewed.

70.   Chan HS, Dill KA: **Transition states and folding dynamics of proteins and heteropolymers.** *J Chem Phys* 1994, 100:9238-9257.

71.   Dill KA, Chan HS: **From Levinthal to pathways to funnels.** *Nat*
••   *Struct Biol* 1997, 4:10-19.
The evolution of theoretical views of folding mechanisms is reviewed, beginning from the seminal ideas of Levinthal and ending with current speculations regarding a new view based on funnel-like energy landscapes. It is argued that the folding pathway concept should be replaced by a more general picture in which proteins fold by trickling down this funnel.

72.   Eyring H. Eyring EM: *Modern Chemical Kinetics.* New York: Reinhold Press; 1963.

73.   Lifshits EM, Pitaevskii LP: *Physical Kinetics.* Oxford: Pergamon Press; 1981.

74.   Kim PS, Baldwin RL: **Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding.** *Annu Rev Biochem* 1982, 51:459-489.

75.   Schmidt FX: **Mechanism of folding of ribonuclease A. Slow refolding is a sequential reaction via structural intermediates.** *Biochemistry* 1983, 22:4690-4696.

76.   Matthews CR: **Pathways of protein folding.** *Annu Rev Biochem* 1993, 62:653-683.

77.   Sali A, Shakhnovich E, 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.

78.   Zwanzig R: **Simple model of protein folding kinetics.** *Proc Natl Acad Sci USA* 1995, 92:9801-9804.

79.   Timoshenko EG, Kuznetsov YuA, Dawson KA: **Kinetics of a Gaussian random copolymer as a prototype for protein folding.** *Phys Rev E* 1996, 54:4071-4086.

80.   Ostrovsky B, Bar-Yam Y: **Motion of polymer ends in homopolymer and heteropolymer collapse.** *Biophys J* 1995, 68:1694-1698.

81.   De Gennes PG: **Kinetics of collapse for a flexible coil.** *J Physique Lett* 1985, 46:L639-L642.

82.   Buguin A, Brochard-Wyart F, de Gennes PG: **Collapse of a flexible coil in a poor solvent.** *CR Acad Sci Paris* 1996, 322:741-746.

83.   Du R, Pande VS, Grosberg A Yu, Tanaka T, Shakhnovich EI: **On
••   the transition coordinate for protein folding.** *J Chem Phys* 1998, 108:334-350.
The probability that a given conformation will fold before unfolding (p_{fold}) is introduced as a quantitative measure of the progress of folding. Without knowledge of the reaction coordinate, knowing p_{fold} enables a true calculation of the transition state (where $p_{fold}=1/2$). p_{fold} can also be used to test whether proposed reaction coordinates (such as Q) are appropriate. It was found that Q is not a good reaction coordinate for protein folding.

84.   Fersht AR: **Nucleation mechanisms in protein folding.** *Curr Opin
••   Struct Biol* 1997, 7:3-9.
The general picture of nucleation is reviewed, in conjunction with a discussion of protein engineering methods. Classical nucleation is compared with nucleation-condensation.

85.   Abkevich VI, Gutin AM, Shakhnovich EI: **Specific nucleus as the transition state for protein folding: evidence from the lattice model.** *Biochemistry* 1994, 33:10026-10036.

86.   Shakhnovich EI, Abkevich VI, Ptitsyn O: **Conserved residues and
••   the mechanism of protein folding.** *Nature* 1996, 379:96-98.
The authors test the nucleation-condensation mechanism by comparing conserved residues of sequences homologous to CI2 with those known to be in the CI2 folding nucleus according to protein engineering (Φ analysis). See also [85], which introduces the idea of a specific nucleus.

87.   Doniach S, Bascle J, Garel T, Orland H: **Partially folded states of proteins: characterization by X-ray scattering.** *J Mol Biol* 1995, 254:960-967.

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

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

90.   Karplus M, Weaver D: **Protein-folding dynamics.** *Nature* 1976, 260:404-406.

91.   Fersht AR: **Characterizing transition states in protein folding: an essential step in the puzzle.** *Curr Opin Struct Biol* 1995, 5:79-84.

92.   Finkelstein AV, Badretdinov AY: **Rate of protein folding near the point of thermodynamic equilibrium between the coil and the most stable chain fold.** *Fold Des* 1997, 2:115-121.

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

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

95.   Tsong TY, Baldwin R: **A sequential model of nucleation-dependent protein folding: kinetic studies of ribonuclease A.** *J Mol Biol* 1972, 63:453-469.

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

97.   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.

98.   Matthews JM, Fersht AR: **Exploring the energy surface of protein folding by structure-reactivity relationships and engineered proteins: observation of Hammond behavior for the gross structure of the transition state and anti-Hammond behavior for structural elements for unfolding/folding of barnase.** *Biochemistry* 1995, 34:6805-6814.

99.   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.

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

101.   Huang GS, Oas TG: **Sub-millisecond folding of monomeric lambda repressor.** *Proc Natl Acad Sci USA* 1995, 92:6878-6882.

102.   Schindler T, Harrier M, Marahiel MA, Schmidt FX: **Extremely rapid protein folding in the absence of intermediates.** *Nat Struct Biol* 1995, 2:663-673.

103.   Ptitsyn OB: **Kinetic and equilibrium intermediates in protein folding.** *Protein Eng* 1994, 7:593-596.

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

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

106.   Raschke TM, Marqusee S: **The kinetic folding intermediate of ribonuclease H resembles the acid molten globule and partially unfolded molecules detected under native conditions.** *Nat Struct Biol* 1997, 4:298-304.

107.   Creighton TE, Darby NJ, Kemmink J: **The roles of partly folded intermediates in protein folding.** *FASEB J* 1996, 10:110-118.

108.   Mirny LA, Abkevich V, Shakhnovich EI: **Universality and diversity of the protein folding scenarios: a comprehensive analysis with the aid of a lattice model.** *Fold Des* 1996, 1:103-116.

109.   Bai Y, Karimi A, Dyson HJ, Wright PE: **Absence of a stable intermediate on the folding pathway of protein A.** *Protein Sci* 1997, 6:1449-1457.

110.   Baldwin RL: **Matching speed and stability.** *Nature* 1995, 369:183-184.

111.   Baldwin RL: **The nature of protein folding pathways: the classical versus the new view.** *J Biomol NMR* 1995, 5:103-109.

112.   Levinthal C: **Are there pathways for protein folding?** *J Chem Phys* 1968, 65:44-45.

113.   Levinthal C: **How to fold graciously.** In *Mossbauer Spectroscopy of Biological Systems.* Edited by Debrunner P, Tsibris JC, Munck E. Urbana, IL: University of Illinois Press; 1969:22-24.

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