Pathways for protein folding: is a "new view" needed?

Vijay S Pande[1], Alexander Yu Grosberg[2], Toyoichi Tanaka[2], and Daniel S Rokhsar[1,3]

Theoretical studies using simplified models for 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.

(December 2, 1997)

Addresses

[1]Department of Physics, University of California at Berkeley, Berkeley, California 94720-7300, USA

[2]Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[3]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

Current Opinion in Structural Biology 1998, **8**: in press

Abbreviations

TS	Transition state
CI2	chymotrypsin inhibitor 2
Q	Number of native contacts
$\mathbf{p}_{\rm fold}$	Folding probability
$\mathbf{R}_{\rm g}$	Radius of gyration
$\mathbf{F}_{\rm int}$	"Internal" free energy
MG	Molten globule

Introduction

How do proteins fold? While the thirty five years since Anfinsen has demonstrated the complexity of protein folding, the search continues for the general principles by which proteins achieve 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 special case of a more general polymeric phenomenon, and much could be learned about the generic aspects of protein folding mechanisms by studying the spontaneous folding of similar polymers. These 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 derived from simplified models of the folding process, considering lattice models for "designed heteropolymers" (defined below) [1–19], simplified models for real proteins [20–28] and all-atom molecular dynamics [29–34] studies. 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 models and 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 for 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, 36–41]. But this "freezing" transition differs from the folding transition in proteins: the folding of random sequences is only weakly cooperative [19,37,38], and proceeds very slowly due to trapping in metastable conformations that are unrelated to the lowest energy conformation [35, 37, 41–43]. Also unlike proteins, the "native" state of a random sequence is 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 desired properties, such as fast folding to a stable, preselected native conformation. In this sense, *design makes heteropolymers protein-like*.[1] 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, 49] and experimentally [50–55].

One design strategy starts with a collection of random sequences and selects only those which fold in a protein-like manner. One can then identify the characteristics of these foldable sequences. This approach has been successful both theoretically [57] and experimentally [52]. In ref. [57], the folding of random sequences was simulated *in silico*, and 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 [52], 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.

[1]Of course, proteins have other characteristics besides folding – such as specific secondary structures – that may not be well-modeled by an overly simplified lattice polymer.

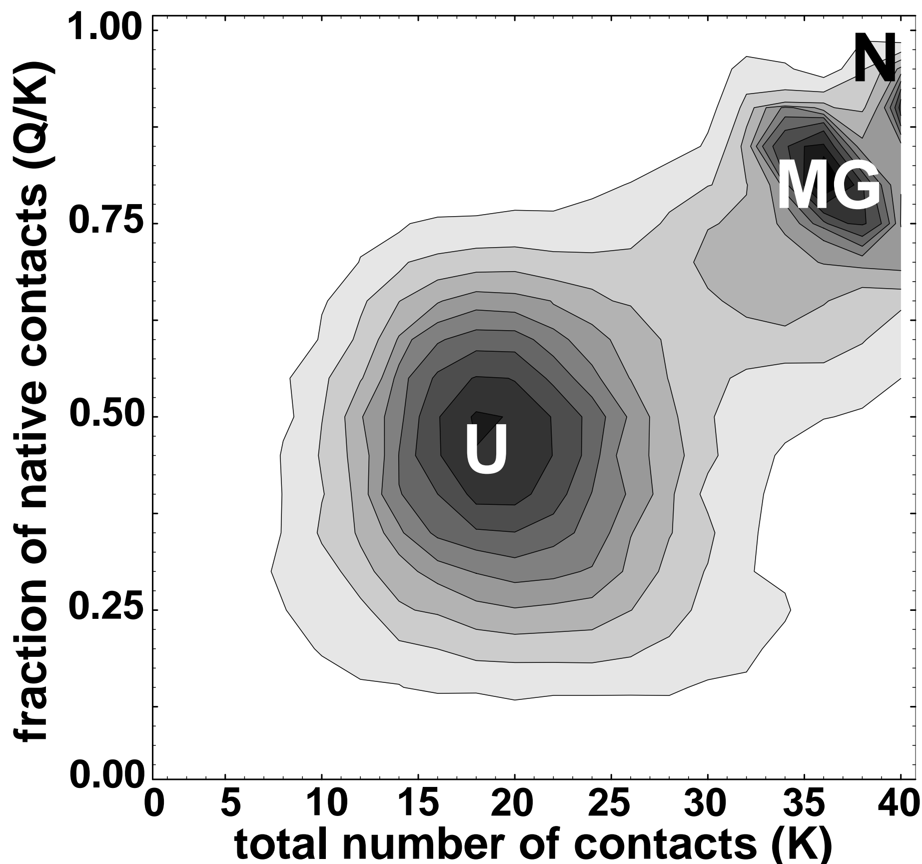


Figure 1: **Free energy landscape.** Free energy contours of a lattice 36-mer calculated from a Monte Carlo simulation [28]. The macroscopic order parameters are the total number of contacts K (both native and non-native) and the fraction of those contacts that are native, Q/K. The three minima correspond to the unfolded (U), molten globule (MG), and native (N) phases, respectively; the intervening barriers imply first order (*i.e.,* 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].

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 relatively low "energy" (*i.e.,* "internal" free energy[2]) in the desired native conformation. This approach has proven successful both in lattice models (reviewed in [12, 19]) and for real proteins (reviewed in [55]). In 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, 48] 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 a designed heteropolymer? These "phases" – the "denatured state," "native state," "molten globule state" [58, 59], *etc.* – correspond to ensembles of conformations that rapidly interconvert on a time scale (picoseconds) much faster than the typical time scale for folding (milliseconds or longer) (see *e.g.*, [60]). The number and nature of the conformations varies for each state; *e.g.,* the denatured state U is associated with an enormous number of largely unrelated unfolded conformations, while the native state N is associated with a few, closely related, low energy conformations.

Order parameters

A useful way to display and conceptualize the phases of a system is to study the free energy as a function of one or more "order parameters," *i.e.,* 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 be within 7Å.) Evidently Q is a good "order parameter" in the sense that it distinguishes the unfolded and folded states: unfolded states typically have 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, $F_{\rm tot}(Q) = F_{\rm int}(Q) - TS_{\rm conf}(Q)$, where $F_{\rm int}(Q)$ is the average "internal" free energy[2] of conformations with Q native con-

[2]The "internal" free energy of a conformation $F_{\rm int}(\mathcal{C})$ – often simply referred to as the "energy" in statistical mechanical models – includes the enthalpy of the polymer as well as the free energy of 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].)

tacts, and $S_{\rm conf}(Q)$ is the corresponding conformational entropy (roughly the logarithm of the number of accessible conformations with Q native contacts). $S_{\rm conf}(Q)$ is easily computed using Monte Carlo simulations, and is close to zero for the native state, but large for the unfolded state.

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

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

(2) Stability: Since different phases respond differently to changes in external conditions, the local minima will shift with temperature, solvent quality, pH, *etc.* If the interconversion time of conformations within a minimum is fast compared with the transition rate to other minima, we may view a relatively high free energy minimum as a "metastable" phase. (Such metastable phases are familiar from the case of a supercooled liquid or gas.) One can sometimes make a particular local minimum *globally* stable by appropriate choice of external conditions. A good example of this is the stabilization of the molten globule 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) Transitions: The free energy barrier between two minima signals 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 described by smooth shifts in the locations of a single minimum with changing external conditions, or the splitting of one minimum into two.

Funnels

It is important to emphasize the distinction between total free energy surfaces and heuristic "funnel" pictures pioneered by Onuchic, Wolynes, Thirumalai *et al.* [8, 43, 63–69], and Chan and Dill [17, 70, 71]. Funnel diagrams plot the *internal* free energy[2] $F_{\rm int}$ (rather than the *total* free energy) *vs.* unspecified "conformation coordinates," and thus do a good job of depicting the "energetic" (really $F_{\rm int}$) 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, cooperativity of the transition, and relative stability of the phases are obscured by the funnel visualization, which does not display entropic barriers (only barriers in the "energy" $F_{\rm int}$). 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 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" (TS). 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 (*i.e.*, 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 $F_{\rm int}$ 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 [57], and can only be characterized statistically. Unlike a simple chemical reaction, in which the free energy barrier represents the contribution from a unique conformation, the barrier for protein folding is a *total* free energy barrier, and may be dominated by conformational entropy (see for example [18]). Such an entropically generated barrier can be thought of as arising from the relative scarcity of TS conformations compared with the unfolded state. Thus, the TS has many conformations compared with the native state, but much fewer than the unfolded state.

Since proteins have many degrees of freedom, in principle 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 slow (preferably the slowest) degree of freedom [72]. The transition state then corresponds to a *total* free energy maximum along the reaction coordinate, rather than an *internal* free energy maximum as would be the case for a simple chemical reaction. Early theoretical work used the number of native contacts Q as a first guess at a reaction coordinate for folding [8, 18, 57, 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.

Protein folding as a first order phase transition

An even closer analogy may be drawn between the folding of a polypeptide chain to its unique native conformation and the transformation of a vapor into a liquid. In both cases, there is a dra-

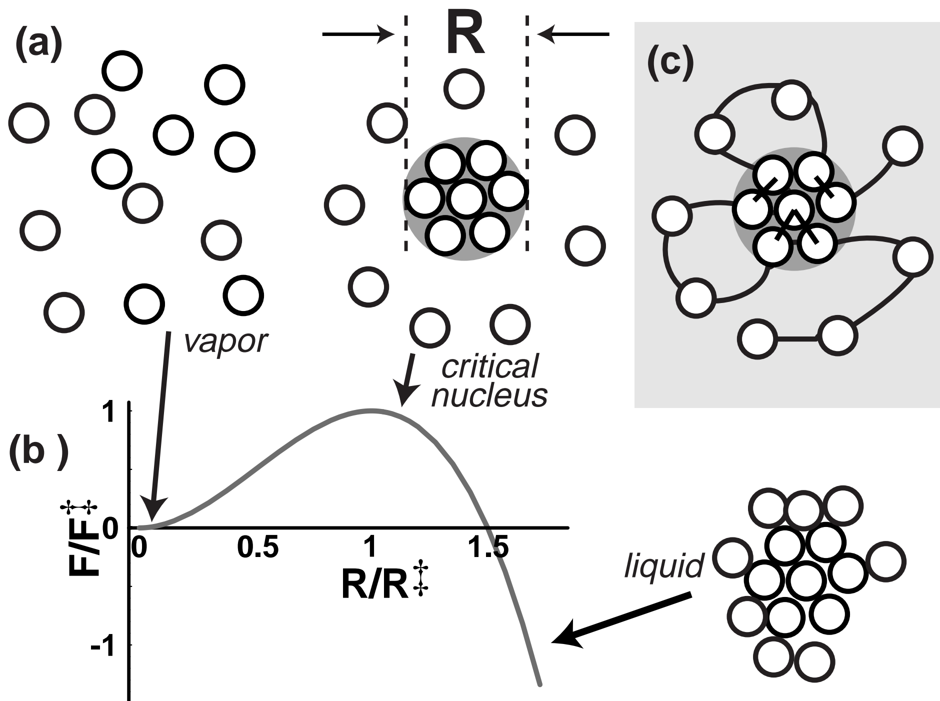


Figure 2: **Nucleation.** **(a)** First order (*i.e.,* cooperative) phase transitions proceed by a nucleation mechanism in which a small droplet of the ordered phase is formed within the metastable disordered phase. **(b)** The free energy of a droplet depends on its radius R. There is a free energy gain proportional to the volume of the droplet, $\sim -\delta f R^3$, where δf is the free energy difference per unit volume between the ordered and disordered phases. Opposing this free energy gain is the cost $\sim \gamma R^2$ of the surface of the droplet, which is the product of the surface tension γ (*i.e.,* 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.

matic 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, *i.e.,* highly cooperative, bulk phase transitions typically proceed *via* a "nucleation and growth" mechanism [73]. Consider a disordered phase (*e.g.,* gas) which is suddenly quenched to a low temperature at which a more ordered phase (*e.g.,* 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" (*i.e.,* a droplet of critical radius $R^{\ddagger}$) is formed (Fig. 2, and caption), 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-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 search for the unique native conformation. The simple homogeneous nucleation picture of a bulk transition will of course need to be modified to account for polymeric chain connectivity [79–83] 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, *i.e.,* a chemical reaction characterized by a single kinetic phase, then we expect to find a well-defined transition state ensemble. This ensemble could be characterized easily if we knew the appropriate reaction coordinate for folding. But from the theory of first-order phase transitions, we expect the rate-limiting step of protein folding to be the formation of some structure or structures analogous to a critical nucleus, which plays the role of a "transition state" for a first order transition.

The physical picture of a first order transition demonstrates that an order parameter for an *equilibrium* transition (*e.g.,* 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 state-

ment applies to the radius of gyration R_g.) Yet from the study of the liquid-gas transition we see that the spatial distribution of the ordered phase (*i.e.,* as droplets) within the disordered state is a central aspect of the mechanism. We therefore expect that the net amount of "order" (*i.e,* 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 likelihood 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.* [84] 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 TS ensemble [85]. They introduce the "folding probability," $p_{\text{fold}}(\mathcal{C})$, which measures the probability that a simulation that starts from conformation $\mathcal{C}$ will reach the folded state before encountering an unfolded conformation. If $\mathcal{C}$ is very close to the native conformation, then $p_{\text{fold}} \approx 1$; if $\mathcal{C}$ is near the unfolded phase, then $p_{\text{fold}} \approx 0$.

The transition state ensemble consists of those conformations sampled during a folding run that have $p_{\text{fold}} = 1/2$, *i.e.,* are equally likely to fold or unfold [3] The relative weight of a conformation in the TS ensemble is defined by its rate of appearance in folding events. In general, conformations with $p_{\text{fold}} = 1/2$ do not appear with equal weight in the transition state ensemble [85]. The folding probability 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

While it is widely believed that some sort of nucleation event is central to the mechanism of protein folding, [23, 39, 86–98], 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 (*i.e.,* how the transition state ensemble is determined) and the resulting picture describing 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 TS ensemble contains a different locally structured region, or "nucleus," reminiscent of the "jigsaw model" of Harrison and Durbin [99]. This picture is supported by the work of Onuchic, Socci, Luthey-Shulten, and Wolynes [8, 69], who investigated the transition state ensemble in 27-mers. They used the number of native contacts Q as a reaction coordinate, computed the total free energy $F_{\text{tot}}(Q)$, and identified the Q_{barrier} at which $F_{\text{tot}}(Q)$ has a maximum. If Q were a good reaction coordinate, conformations with $Q = Q_{\text{barrier}}$ would comprise the transition state ensemble. The fact that this analysis is based on the flawed 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 is comprised of many partially folded conformations. A 27-mer has a total of 10^{16} possible conformations, with 10^{10} of them "semi-compact," *i.e.,* highly collapsed [57]. Onuchic *et al.* 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 TS 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 term "delocalized."

In an earlier study, Sali, Shakhnovich, and Karplus [57] also examined the folding of designed 27-mers. Also using Q as a reaction coordinate, they identified a different barrier in $F_{\text{tot}}(Q)$ as the major transition state, and inferred that the TS ensemble consists of all 10^3 semi-compact 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 [100] as evidence for many parallel pathways of folding.

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]. They introduce a "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, and conclude that the TS ensemble is "49% of the way from the native state to the 'unfolded state' " [17].

Evidence for a "specific nucleus"

A qualitatively different kind of transition state ensemble was proposed by Abkevich, Gutin, and Shakhnovich [87] based on their analysis of the folding of a designed 36-mer. 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 is comprised of conformations that share the same set of essential contacts, which form a compact core inside the native state – a "specific nucleus" [87–89]. As a test of this hypothesis, Shakhnovich, Abkevich, and Ptitsyn [89] have confirmed that different sequences designed for the CI2 backbone have conserved residues at the predicted core positions. This picture closely resembles nucleation in first order phase transitions, with a critical

[3] For a two-state transition, there is a single, well-defined transition state. If there are intermediates, then $p_{\text{fold}} \approx 1/2$ determines the "major" transition state that governs the rate limiting step.

nucleus pinned 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 TS ensemble, then this ensemble would comprise related conformations that differ only in the configuration of polymeric loops that lie between core contacts (Fig 2c). Thus despite the formation of an ordered core, the TS ensemble in the specific nucleus picture has a substantial entropy that arises from the conformational freedom of these loops.

Evidence for "transition state classes"

A third scenario is proposed by Pande and Rokhsar [85], who analysed folding pathways and the transition state ensembles for a range of polymer lengths from 27- to 64-mers. They directly determine the transition state ensemble using the folding probability method, thereby avoiding ambiguities associated with choosing a reaction coordinate. The transition state ensemble is defined by collecting $p_{\text{fold}}(\mathcal{C}) = 1/2$ conformations from several hundred folding trajectories (using the same sequence, but starting from different unfolded conformations). For 27-mers, the TS 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 in the specific nucleus picture, the conformational freedom of loops endows the transition state with a large entropy. However, 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 TS ensemble of fast-folding sequences [88] consists of a single class [85].

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 [85] using the p_{fold} method [84] support the transition state class scenario in lattice models, the nature of the transition state ensemble in real proteins can be best addressed by experiments.

Φ-value analysis

The principal experimental method for identifying transition states for folding is the Φ analysis introduced by Fersht *et al.* [86, 94]. They use site-directed mutagenesis 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. (Here Δ refers to the difference between the mutant and wild type proteins.) A residue which 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 (*i.e.,* between 0.3 and 0.6), which can be interpreted in two ways: (a) the residue makes native-like contacts in only a fraction of the transition state conformations, or (b) the residue makes contacts in the transition state ensemble that are weakened relative to those it forms in the native state. Fersht *et al.* favor (b) based on a comparison of single *vs.* multiple pathway models with kinetic data [100, 101].

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. The histogram [8] of experimentally determined Φ values for CI2 [94, 102] is broadly peaked between $\Phi = 0$ and 0.6. Onuchic *et al.* [8] find a similarly broad distribution of Φ-value analogs for a lattice 27-mer (which they argue is comparable to a 60-residue protein). The molecular dynamics sampling of fragment B of 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 a strict "specific nucleus" picture were valid, the Φ value probability distribution would be bimodal – residues in the nucleus would have high Φ, while residues not in the nucleus should have $\Phi \sim 0$.

The simulations of Pande and Rokhsar [85] also reproduce a broad distribution of Φ-value analogs (the fraction of TS conformations that possess a given native contact). Unlike Onuchic *et al.*, however, they explain the broad distribution of Φ-values by appealing to 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.* rule out multiple pathways for CI2 by appealing to a Bronsted analysis [100], in which the logarithm of the folding (or unfolding) rate is plotted *vs.* the detabilization 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 [101].

Molecular dynamics simulations of unfolding at high temperature

All-atom simulations of unfolding trajectories of chymotrypsin inhibitor II (CI2) under extreme conditions (500 K, 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 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 worry that the TS 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

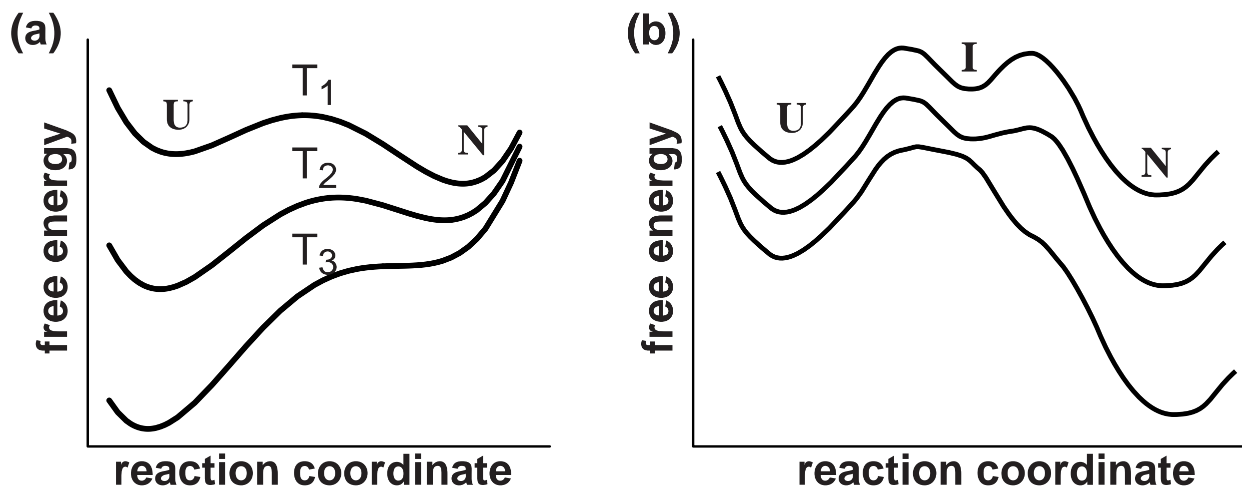


Figure 3: **Temperature dependence of barriers and intermediates.** (a) This schematic plot of total free energy *vs.* a generic "reaction coordinate" for several temperatures illustrates that barriers are temperature dependent. Under extreme conditions ($T_3 \gg T_2$), the free energy barrier may disappear.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

be present under extreme temperature and pressure if the native state loses its metastability (see Fig. 3a). Nevertheless, there is remarkable consistency between the residues Daggett *et al.* identify as important in the transition state and those implicated experimentally by Fersht *et al.* using Φ-value analysis.

Designing pathways

A complete understanding of the mechanism(s) of protein folding should include a prescription for redesigning folding pathways. That is, in addition to designing *equilibrium* properties of a heteropolymer, one should be able to intentionally manipulate its folding *kinetics*. As a first step in this direction, Shakhnovich and his collaborators have used an evolution-like process to select fast folding lattice heteropolymer sequences by mutating sequences and retaining those variants that fold most quickly [88].

Mirny *et al.* [88] find that all fast folding sequences designed in this manner fold with the same specific nucleus. An analysis of the pathways of these fast-folding sequences by Pande and Rokhsar [85] using the $p_{\rm fold}$ method shows that they fold *via* a single "transition state class" that is energetically preferred from among several possible classes found in typical heteropolymers designed for equilibrium folding to that native state conformation. That is, evolutionary design for fast folding leads to a specific pathway. Pande and Rokhsar ([85] and unpublished results) have used this idea to directly design sequences (that is, without an evolutionary selection for fast folding) with both (a) a preselected native state conformation and (b) 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 contradicts the "many delocalized nuclei" scenario.

Intermediates

Many small proteins fold without detectable intermediates [91, 103–105]. Yet there are clear examples of others whose folding route passes through partially folded, molten-globule-like, "on-pathway" intermediates [61, 106–109]. Still 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 [110]. Such intermediates are either inferred from multi-state 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 found such intermediates, which may be due to differences in the methodologies of the different calculations (for example, different temperatures of the simulations), or 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 (*i.e.,* locally stable minima of the free energy surface (Fig. 1)), then as folding temperature, pressure, pH, *etc.* are varied the stability of such a state will change and may disappear (Fig. 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 [85] found that each transition state class is associated with a corresponding on-pathway, partially folded intermediate. The conformations which 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 $Q_{\rm core}$; $F(Q, Q_{\rm core})$ exhibits a metastable intermediate minimum along with the unfolded and folded minima. At sufficiently low temperatures the barrier between the intermediate and native states disappears, and the transition becomes two-state

(Fig. 3).

Mirny, Abkevich, and Shakhnovich [111] find that well-designed sequences are more stable in the native state and fold fast without intermediates in a two-state process, whereas less-optimized sequences fold more slowly, *via* parallel pathways involving misfolded intermediates.

Off-pathway intermediates have been found in the coarse-grained, non-lattice models of four helix bundles studied by Thirumalai *et al.* [26, 43]. They perform Langevin dynamics simulations in which the polypeptide is 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 find intermediates that are misfolded (one of the helices is 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] have 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 300 K to 400 K), and sampled conformations at many values of the radius of gyration $R_{\rm g}$ to piece together the free energy $G(R_{\rm g})$. Conformations generated in this run are used to construct "clusters" with given $R_{\rm g}$. From this analysis, they infer a folding intermediate for this small protein. Recent experiments on protein A [112], 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 (*i.e.,* partly folded) or off-pathway (*i.e.,* misfolded) intermediates [74, 75]. Several years ago, Baldwin [113, 114] suggested that a "new" view was emerging based on simplified statistical mechanical models for proteins. As we have seen, these models emphasize 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 the new view is the replacement of the pathway concept with picturesque funnel diagrams to illustrate features of protein folding and the role of ensembles. This "funnelist" viewpoint has recently been reviewed in detail by Chan and Dill [71].

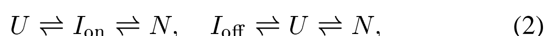
In contrast, other studies of simple models [56, 85] do not suggest a radical "new" view, but rather a refinement of the classical picture in which the classical concepts of "states" and "pathways" are interpreted in terms of ensembles of conformations. For example, each step in a classical pathway can be precisely regarded as a transition between two phases (*i.e.*, ensembles of rapidly interconverting conformations [60]), so that folding proceeds through a sequence of metastable phases. In the next section, we briefly summarize this alternative to the funnel picture, which might be called a "neo-classical" view.

"Classical" pathways from an ensemble view

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

$$U \rightleftharpoons N, \qquad (1)$$

models with on- or off-pathway intermediates,



$$U \rightleftharpoons I_{\rm on} \rightleftharpoons N, \quad I_{\rm off} \rightleftharpoons U \rightleftharpoons N, \qquad (2)$$

etc. Increasingly complex schemes become increasingly difficult to compare with experiment, and there are as yet no first principles rules to determine in advance which pathway will apply to a specific protein.

While in simple chemical reactions the symbols in the "mass action" equations (1) and (2) represent *specific* conformations of a small molecule, for protein folding we must interpret each symbol as an *ensemble* of rapidly interconverting conformations, *i.e.,* a thermodynamic phase. In a temperature jump experiment, for example, U would be a supercooled, (*i.e.,* 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, molten-globule-like phases [28, 85]; others find misfolded, *i.e.,* off-pathway, intermediates [26, 111].

In this ensemble view of a classical pathway, the "$\rightleftharpoons$" arrows denote first order, *i.e.,* 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 than the funnelist viewpoint, which replaces the chemical reaction analogy with a picture of 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 [115–117]. While this is true, it is also irrelevant: the randomly fluctuating unfolded polypeptide only needs to "find" one of the many member 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 [57]: $t_f^{\mathrm{random}} = t_0(W_{\mathrm{UF}}/W_{\mathrm{TS}})$, *i.e.*, the typical time to sample a distinct conformation (t_0) multiplied by the ratio of the number of unfolded states (W_{UF}) to the number of transition states (W_{TS}).

Using Q as a reaction coordinate to describe the transition state, Sali, Shakhnovich, and Karplus [57, 77] suggested that in 27-mer lattice models, the polymer "finds" a member of the transition state ensemble by random search. Using the more reliable p_{fold} approach [84] and longer chains, Pande and Rokhsar [85] have demonstrated the existence of a random search mechanism using three 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 from Monte Carlo simulations agree with the calculation of t_f^{random} employing simulation measurements of W_{UF} and W_{TS}. (The combinatorial entropy of the optional TS contacts is critical for this agreement.) Finally, the internal free energies of conformations in the transition were comparable to the internal free energy of the unfolded state, indicating that there is not a substantial internal free energetic drive towards the transition state.[4]

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, *i.e.,* the "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 the 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" [115], where "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 "neoclassical" view develops, we look for increasing comparisons with experiments, the ultimate arbiter of theoretical progress.

Acknowledgements

VSP and DSR thank the Miller Institute for Basic Research in Science for support. AYG and TT acknowledge support from National Science Foundation DMR 90-22933 and MCB-93-16186. DSR acknowledges support from the National Science Foundation DMR-91-57414, Lawrence Berkeley National Laboratory Grant LDRD-3669-57, and the National Energy Research Scientific Computing Center. We gratefully thank Dan Butts, Arup Chakraborty, Aaron Chamberlain, David Chandler, Rosy Cho, Tracy Handel, Susan Marqusee, Jennifer Nickel, and Tanya Raschke for critical readings of the manuscript and helpful discussions.

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**:445-59.
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 Nat Acad Sci USA* 1993, **90**:7195-9.
6. Camacho CJ Thirumalai D: **Kinetics and thermodynamics of folding in model proteins**. *Proc. Natl. Acad. Sci. USA* 1993, **90**:6369-72.
7. Pande VS, Grosberg AYu, Tanaka T: **Thermodynamic Procedure to Synthesize Heteropolymers that can Renature to Recognize a Given Target Molecule.** *Proc Nat Acad Sci USA* 1994, **91**:12976-81.
8. • Onuchic JN, Socci ND, Luthey-Schulten Z, Wolynes PG: **Protein folding funnels: the nature of the transition state ensemble.** *Fold Des* 1996, **1**:441-50.
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** *Prot Struct Funct Genet* 1996, **24**:335-344.
10. Hinds DA, Levitt M: **From structure to sequence and back again.** *J Mol Biol* 1996, **258**:201-9.

[4]This does not, of course, imply the absence of an activation enthalpy, since the internal free energy of a conformation includes the entropy of the surrounding solvent.[2]

11. Abkevich VI, Gutin AM, Shakhnovich EI: **Improved design of stable and fast-folding model proteins.** *Fold des* 1996, **1**:221-30.
12. • Shakhnovich EI: **Modeling protein folding: the beauty and power of simplicity.** *Folding and Design* 1996, **1**:R50-52.
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**:286-99.
15. Kurosky T, Deutsch JM: **Design of Copolymeric Materials** J Phys A 1996, **27** L387-9.
16. Deutsch JM, Kurosky T: **New Algorithm for Protein Design** Phys Rev Let 1996, **76**, 323-6.
17. Chan HS, Dill K: **Protein folding kinetics from the perspective of simple models.** *Prot Struct Funct Gen* 1997, **in press**.
18. Pande VS, Grosberg AY, Tanaka T: **On the theory of folding kinetics for short proteins.** *Fold Des* 1997, **2**:109-114.
19. Pande VS, Grosberg AYu, Tanaka T: **Heteropolymer Freezing and Design: Towards Physical Models of Protein Folding.** *Reviews of Modern Physics* 1998, **in press**.
20. Levitt M, Warshel A: **Computer simulation of protein folding.** *Nature* 1975, **253**:694-8.
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**:271-87.
25. • Guo Z, Thirumalai D: **Kinetics and thermodynamics of folding of a de novo designed four-helix bundle protein.** *J Mol Biol* 1996, **263**:323-43.
A simplified representation of a polypeptide is introduced to simulate the folding kinetics of a four helix bundle. Misfolded intermediates are identified.
26. Camacho CJ, Thirumalai D: **Denaturants can accelerate folding rates in a class of globular proteins.** *Prot Sci* 1996, **5**:1826-32.
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 Nat Acad Sci USA* 1998, in press.
The molten globule phase is found in lattice and off lattice protein models and it is suggested that this phase is a general heteropolymer property. The stability of the molten globule is discussed in terms of conformational entropy of disordered loops.
29. Caflisch A, Karplus M: **Molecular dynamics simulation of protein denaturation: solvation of the hyrophobic core and secondary structure of barnase.** *Proc Nat 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 Nat Acad Sci USA* 1994, **91**:10430-4.
All-atom molecular dynamics simulations of unfolding under extreme conditions (500 K, 26 atmospheres) are used to identify the transition state for unfolding. Although the trajectories are quite short (a few nanoseconds) the inferred transition state is consistent with that found by Fersht *et al.* using the protein engineering method.
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**:501-20.
32. Van Gunsteren WF, Hunenberger PH, Kovacs H, Mark AE, Schiffer CA. **Investigation of protein unfolding and stability by computer simulation.** *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 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**:393-6.
Thermodynamics properties of an all atom model of fragment B of protein A are calculated by the technique of molecular dynamics sampling, and a proposed free energy surface *vs.* the radius of gyration is obtained.
34. Tirado-Rives J, Orozco M, Jorgensen WL: **Molecular dynamics simulations of the unfolding of barnase in water and 8 M aqueous urea.** *Biochem* 1997, **36**:7313-29.
35. 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.
36. Bryngelson JD, Wolynes PG: **Spin glasses and the statistical mechanics of protein folding.** *Proc Nat Acad Sci USA* 1987, **84**:7524-8.
37. Sfatos CD, Gutin AM, Shakhnovich EI: **Phase diagram of random copolymers.** *Phys Rev E* 1993, **48**:465-75.
38. Pande VS, Grosberg AY, Tanaka T: **Freezing Transition of Random Heteropolymers Consisting of an Arbitrary Set of Monomers.** *Phys Rev* 1995, **E51**, 3381-3402.
39. Gutin AM, Abkevich VI, Shakhnovich EI: **Chain length scaling of protein folding time.** *Phys Rev Lett* 1996, **77**:5433-5436.
The scaling of characteristic folding times with polymer length is discussed for lattice heteropolymers of between 27 and 200 residues. While the Levinthal time scales exponentially with length, the observed power law scaling supports a nucleation scenario.
40. 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.
41. Wang J, Plotkin SS, Wolynes PG: **Configurational diffusion on a locally connected correlated energy landscape: application to finite, random heteropolymers.** *J de Phys I (France)* 1997, **7**:395-421.
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.
43. •• Wolynes PG, Onuchic JN, Thirumalai D: **Navigating the folding routes.** *Science* 1995, **267**:1619-20.
Protein folding is described 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.
44. Shakhnovich EI, Gutin AM: **Influence of point mutations on protein structure: probability of a neutral mutation.** *J Theor Biol* 1991, **149**: 537-46.
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**:6038-45.
46. Pande VS, Grosberg AY, Tanaka T: **How accurate must potentials be for successful modeling of protein folding?** *J Chem Phys* 1995, **103**:9482.
47. Bussemaker HJ, Thirumali D, Bhattacherjee: **Thermodynamic**

stability of folded proteins against mutations. *Phys Rev Lett* 1997, **79**:3530-3.
48. Ramanathan S, Shakhnovich E: **Statistical mechanics of proteins with "evolutionary selected" sequences.** *Phys Rev* 1994, **E50**:1303-12.
49. • Shakhnovich EI: **Proteins with selected sequences fold into unique native conformation.** *Phys Rev Lett* 1994, **72**:3907-3910.
50. Regan L, DeGrado WR: **Characterization of a helical protein designed from first principles.** *Science* 1988, **241**:976-978.
51. 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.
52. Davidson AR, Lumb KJ, Sauer RT: **Cooperatively folded proteins in random sequence libraries.** *Nature Struct Biol* 1995, **2**:856-863.
53. Dahiyat BI, Mayo SL: **Protein design automation.** *Protein Science* 1996, **5**:895-903.
54. Lazar GA, Desjarlais JR, Handel TM. **De novo design of the hydrophobic core of ubiquitin**. *Protein Science* 1997, **6**:1167-1178.
55. DeGrado WF: **Proteins from scratch.** *Science* 1997, **278**:80-1.
56. Shakhnovich EI: **Theoretical studies of protein-folding thermodynamics and kinetics.** *Curr Opin Struct Biol* 1997, **7**:29-40.
57. •• Sali A, Shakhnovich EI, Karplus M: **How does a protein fold?** *Nature* 1994, **369**:248-51.
The folding of 27-mer lattice heteropolymers is described in detail. The transition state ensemble is calculated assuming that the number of native contacts Q is an appropriate reaction coordinate, and the "random search for the transition state ensemble" mechanism is proposed.
58. Ptitsyn OB, Uversky VN: **The molten globule is a third thermodynamical state of protein molecules.** *FEBS Lett* 1994, **341**:15-8.
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**:117-22.
60. Zwanzig R: **Two-state models of protein folding kinetics.** *Proc Nat Acad Sci, USA* 1997, **94**:148-50.
61. Ptitsyn OB: **Structures of folding intermediates.** *Curr Op Struct Biol* 1995, **5**:74-8.
62. Haynie DT, Freire E: **Structural energetics of the molten globule state.** *Proteins* 1993, **16**:115-40.
63. Leopold PE, Montal M, Onuchic, JN: **Protein folding funnels: a kinetic approach to the sequence-structure relationship.** *Proc Nat Acad Sci USA* 1992, **89**:8721-5.
64. Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG: **Funnels, pathways, and the energy landscape of protein folding: A synthesis**. *Proteins: Struct Funct Genet* 1995, **21**:167-195.
65. Onuchic JN, Wolynes PG, Luthey-Shulten Z, Socci ND: **Towards and outline of the topography of a realistic protein folding funnel.** *Proc Nat Acad Sci, USA* 1995, **92**:3626-3630.
66. Shoemaker BA, Wang J, Wolynes PG: **Structural correlations in protein folding funnels.** *Proc Nat Acad Sci* 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.
70. Chan HS, Dill KA: **Transition states and folding dynamics of proteins and heteropolymers.** *J Chem Phys* 1994, **100**:9238-57.
71. •• Dill KA, Chan HS: **From Levinthal to pathways to funnels.** *Nat Struct Biol* 1997, **4**:10-9.
The evolution of theoretical views of folding mechanism 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 with 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. Schmid 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 Nat Acad Sci* 1995, **92**:9801-4.
79. Garel T, Orland H, Pitard E: **Protein folding and heteropolymers.** Preprint cond-mat 9706125 June 12, 1997.
The authors present analytical arguments regarding the initial stage of folding, which resembles a homopolymer collapse. A useful review of theoretical models of secondary structure formation in the course of collapse is also provided.
80. Timoshenko EG, Kuznetsov YuA, Dawson KA: **Kinetics of a Gaussian random copolymer as a prototype for protein folding.** *Phys Rev E* 1996, **54**:4071-86.
An attempt of the Langevin-type self-consistent theory of a random sequence heteropolymer collapse. See also an alternative view in [81].
81. Ostrovsky B, Bar-Yam Y: **Motion of polymer ends in homopolymer and heteropolymer collapse.** *Biophys J* 1995, **68**:1694-8.
82. de Gennes PG: **Kinetics of collapse for a flexible coil.** *J Physique Lett* 1985, **46**:L639-42.
83. Buguin A, Brochard-Wyart F, de Gennes PG: **Collapse of a flexible coil in a poor solvent.** *CR Acad Sci Paris* 1996, **322 IIb**:741-6.
A revision of the homopolymer collapse theory [82].
84. •• Du R, Pande VS, Grosberg AY, Tanaka T, Shakhnovich EI: **On the transition coordinate for protein folding.** *J Chem Phys* 1998 **in press**.
The probability that a given conformation will fold before unfolding (p_{fold}) is introduced as a quantitative measure of folding progress. Without knowledge of the reaction coordinate, p_{fold} allows for a true calculation of the transition state (where $p_{\text{fold}} = 1/2$). Also one can use p_{fold} to test whether proposed reaction coordinates (such as the number of native contacts Q) are appropriate. It is found that Q is not a good reaction coordinate for protein folding.
85. •• Pande VS, Rokhsar DS: **Direct approach to protein folding pathways.** submitted to *Nat Struct Biol* 1997.
The nature of the transition state and the folding pathway is quantitatively analyzed for a lattice heteropolymer model. The transition state ensemble is calculated using the technique outlined in Du, *et al.* [84], and is shown to consist of many, highly related confor-

mations – a “transition state class.” On-pathway intermediates that resemble molten globule-like phases are identified both kinetically and as metastable equilbrium phases.

86. •• Fersht AR: **Nucleation mechanisms in protein folding.** *Curr Op Struct Biol* 1997. **7**:3-9.
The general picture of nucleation is reviewed, in conjunction with a discussion of the protein engineering method. Classical nucleation is compared with nucleation-condensation.
87. •• Abkevich VI, Gutin AM, Shakhnovich EI: **Specific nucleus as the transition state for protein folding: evidence from the lattice model.** *Biochem* 1994, **33**:10026-36.
Simulation data is presented to support the idea that the transition state is an ensemble of conformations that contain a specific set of contacts – a “specific nucleus.” Once these critical contacts have formed, folding is fast.
88. Abkevich VI, Mirny LA, Shakhnovich EI: **Key nucleus positions that determine the structure of a protein: emergence in evolutionary selection and role in protein folding.** *Proc Natl Acad Sci USA* 1997, submitted.
89. Shakhnovich EI, Abkevich VI, Ptitsyn O: **Conserved residues and the mechanism of protein folding.** *Nature* 1996, **379**:96-8.
The authors the “nucleation-condensation” mechanism to infer the residues that participate in the folding nucleus of CI2 by assuming that such residues will be conserved between homologous sequences.
90. Doniach S, Bascle J, Garel T, Orland H: **Partially folded states of proteins: characterization by X-ray scattering.** *J Mol Biol* 1995, **254**:960-7.
91. Fersht AR: **Optimization of rates of protein folding: the nucleation-condensation mechanism and its implications.** *Proc Natl Acad Sci USA* 1995, **92**:10869-73.
92. Wetlaufer DB: **Nucleation, rapid folding, and globular interchain regions in proteins.** *Proc Natl Acad Sci USA* 1973, **70**:697-701.
93. Karplus M, Weaver D: **Protein-folding dynamics.** *Nature* 1976 **260**:404-406.
94. Fersht AR: **Characterizing transition states in protein folding: an essential step in the puzzle.** *Curr Op Struct Biol* 1995, **5**:79-84.
95. 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.
96. Unger R, Moult J: **Local interactions dominate folding in a simple protein model.** *J Mol Bio* 1996, **259**:988-94.
97. Moult J, Unger R: **An analysis of protein folding pathways.** *Biochemistry* 1991, **30**:3816-24.
98. Tsong TY, Baldwin R: **A sequential model of nucleation-dependent protein folding: kinetic studies of ribonuclease A.** *J Mol Biol* 1972, **63**:453-469.
99. Harrison SC, Durbin R: **Is there a single pathway for the folding of a polypeptide chain?** *Proc Nat Acad Sci, USA* 1985, **82**:4028-30.
100. 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 Nat Acad Sci USA* 1994, **91**:10426-10429.
101. 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.** *Biochem* 1995, **34**:6805-6814.
102. 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 Nat Acad Sci USA* 1994, **91**:10422-5.
103. Sosnick TR, Mayne L, Hiller R, Englander SW: **The barriers in protein folding.** *Nature Struct Biol* 1994, **1**:149-156.
104. Huang GS, Oas TG: **Sub-millisecond folding of monomeric lambda repressor.** *Proc Nat Acad Sci, USA* 1995, **92**:6878-6882.
105. Schindler T, Harrier M, Marahiel MA, Schmidt FX: **Extremely rapid protein folding in the absence of intermediates.** *Nat Struct Biol* 1995, **2**:663-673.
106. Ptitsyn OB: **Kinetic and equilibrium intermediates in protein folding.** *Protein Eng* 1994, **7**:593-6.
107. Dobson CM, Evans PA, Radford SE: **Understanding how proteins fold: the lysozyme story so far.** *Trends Biochem Sci* 1994, **19**:31-37.
108. Jennings PA, Wright PE: **Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin.** *Science* 1993, **262**:892-896.
109. 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.
110. Creighton TE, Darby NJ, Kemmink J: **The roles of partly folded intermediates in protein folding.** *Faseb Journal* 1996, **10**:110-8.
111. 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**:103-16.
112. Bai Y, Karimi A, Dyson HJ, Wright PE: **Absence of a stable intermediate on the folding pathway of protein A.** *Prot Sci* 1997, **6**:1449-57.
113. Baldwin RL: **Matching speed and stability.** *Nature* 1995, **369**:183-4.
114. •• Baldwin RL: **The nature of protein folding pathways: the classical versus the new view.** *J Biomol NMR* 1995, **5**:103-9.
Recent experiments are discussed in the context of statistical mechanical models for folding, which emphasize an ensemble view of the folding process. The importance of folding intermediates for fast folding proteins is reexamined, and a “new view” of folding pathways is delineated.
115. Levinthal C: **Are there pathways for protein folding?** *J Chim Phys* 1968, **65**:44-5.
116. Levinthal C: **How to fold graciously.** *Mossbauer Spectroscopy of Biological Systems.* pp. 22-24. eds. Debrunner, P; Tsibris, JC; Munck E. (U. Illinois Press, Urbana IL).
117. Zwanzig R, Szabo A, Bagchi B: **Levinthal's paradox.** *Proc. Natl. Acad. Sci. USA* 1992, **89**:20-22.