PROTEINS: Structure, Function, and Genetics 32:136–158 (1998)

RESEARCH ARTICLES

Protein Folding Mechanisms and the Multidimensional Folding Funnel

Nicholas D. Socci,[1,2]* José Nelson Onuchic,[2] and Peter G. Wolynes[3]
[1]*Bell Laboratories, Lucent Technologies, Murray Hill, New Jersey*
[2]*Department of Physics, University of California at San Diego, La Jolla, California*
[3]*School of Chemical Sciences, University of Illinois, Urbana, Illinois*

ABSTRACT **An important idea that emerges from the energy landscape theory of protein folding is that subtle global features of the protein landscape can profoundly affect the apparent mechanism of folding. The relationship between various characteristic temperatures in the phase diagrams and landmarks in the folding funnel at fixed temperatures can be used to classify different folding behaviors. The one-dimensional picture of a folding funnel classifies folding kinetics into four basic scenarios, depending on the relative location of the thermodynamic barrier and the glass transition as a function of a single-order parameter. However, the folding mechanism may not always be quantitatively described by a single-order parameter. Several other order parameters, such as degree of secondary structure formation, collapse and topological order, are needed to establish the connection between minimalist models and proteins in the laboratory. In this article we describe a simple multidimensional funnel based on two-order parameters that measure the degree of collapse and topological order. The appearance of several different "mechanisms" is illustrated by analyzing lattice models with different potentials and sequences with different degrees of design. In most cases, the two-dimensional analysis leads to a classification of mechanisms totally in keeping with the one-dimensional scheme, but a topologically distinct scenario of fast folding with traps also emerges. The nature of traps depends on the relative location of the glass transition surface and the thermodynamic barrier in the multidimensional funnel. Proteins 32:136–158, 1998.**

Key words: lattice models; landscape theory; Monte Carlo simulations; folding funnels

INTRODUCTION

The beautiful structure of folded proteins as revealed by static structural studies is characterized by several kinds of order. Most dynamical experiments on folding sense only one kind of order at a time. Some probes of folding kinetics determine compactness, others determine secondary structure characteristics, and still others are sensitive to specific tertiary contacts. In general, the folding process can be described by energy landscape theory and the concept of a folding funnel.[1–17] In order to completely understand a set of folding experiments, the characteristics of this funnel must be understood through a multidimensional view of its shape and statistical properties. Earlier theoretical work has shown how a rich variety of folding mechanisms can be understood with a one-dimensional picture of the folding funnel, and thus described by a single-order parameter. In this article we illustrate the way in which a multidimensional view of a folding funnel allows one simultaneously to understand folding kinetics as seen by different probes. Specifically we discuss the way in which different scenarios of folding can arise, depending on the interplay between the collapse-order parameter and an order parameter sensitive to specific tertiary contacts. The two-dimensional folding funnel allows us to classify and compare a variety of observations noted in different simulations[16,18–24] and experiments.[25] For example, one popular mechanistic scheme, the kinetic partitioning model (KPM),[26,27] in which in some cases the chain folds rapidly without trapping and other times is trapped and must unfold to escape, is shown to

Grant sponsor: Bell Laboratories, Lucent Technologies; Grant sponsor: National Science Foundation; Grant number: MCB-9603839; Grant sponsor: University of California; Grant sponsor: National Institutes of Health; Grant number: 1R01 GM44557.
*Correspondence to: Dr. Nicholas D. Socci, Bell Laboratories, Lucent Technologies, 700 Mountain Avenue, Murray Hill, NJ 07974.
Received 9 June 1997; Accepted 24 November 1997

correspond to one class of behavior in the two-dimensional funnel representation (the low-hydrophobicity, low-temperature range of a minimally frustrated sequence). There are some key differences from that phenomenology where, in particular, generally multiexponential behavior occurs in the simulations, while the KPM treats trap escape as exponential. The degree of nonexponentiality is also related to the structure of the landscape. Again, earlier simulation studies have seen many of these behaviors but have not tried to show their quantitative connection to statistical landscape topography. Here we see they depend on the relation of three temperatures characterizing the folding transition, collapse transition, and glass transition.

The one-dimensional (single-order parameter) view of a folding funnel allows the classification of folding into four basic scenarios, depending on the relation between the position of the thermodynamic barrier and of the ideal thermodynamic glass transition (which signals the appearance of discrete kinetic intermediates, i.e., traps) as a function of the order parameter. The simplest situations with multiple-order parameters consist of combinations of these different scenarios for each of the different order parameters. A topologically different situation can arise in the multidimensional case where the character of the dominant flows in the funnel may be coupled, leading sometimes to simultaneity of order formation, or alternatively to a partitioning of folding routes, leading to multiexponential macroscopic kinetics.

The variety of mechanisms and of dominant flows in the two-dimensional folding funnel are illustrated by simulations on a lattice model with a varying tendency to general hydrophobic collapse. The analysis shows how understanding the folding mechanism requires a consideration of the phase diagram of a real protein in order to set up the correspondence between computer models and laboratory experiments. While at the moment the requisite experiments are not in hand to set up such a correspondence completely quantitatively, the dependence of global mechanism on denaturant, as well as other agents affecting other order parameters, for example, a secondary structure, like TFE, can be rationalized by using the multidimensional funnel phenomenology.

The simplest example of a multidimensional folding funnel emphasizing collapse and topological order uses as order parameters the number of contacts as one variable (**Z**) and the number of native contacts (**Q**; that is, the number of correct contacts. **Q** measures the fraction of the distance map of the protein congruent to fully folded structure. According to analytical theory, the coupling between these order parameters can be a major source of the barrier separating native and unfolded states.[28] Also, the glass transition is a strong function of the degree of collapse as noticed by Bryngelson and Wolynes[3] and later workers and confirmed by simulations.[29] These dependencies lead to an interesting set of scenarios. Conjugated to these order parameters are energetic parameters that are crudely tunable by experiments. One energy scale determines the drive to collapsed structures and corresponds to the nonspecific tendency for hydrophobic collapse. The other energetic parameter characterizes the stability gap between the native structure and other compact conformations, giving rise to the specific native structure. The latter parameter depends on the degree of frustration of the sequence being studied. In the laboratory, it can be varied by using site mutations that perturb native stability.

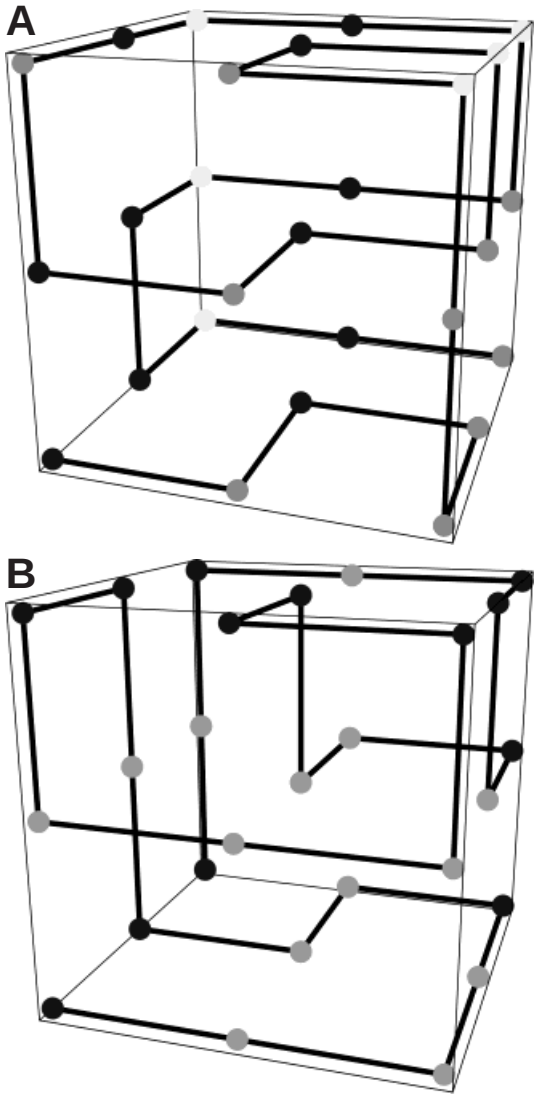

Fig. 1. Two 27 length polymers on a three-dimensional simple cubic lattice. The conformation is a $3 \times 3 \times 3$ maximally compact cube. The different shade spheres represent the different monomer types. **A:** Minimally frustrated (good) sequence (ABABBBC-BACBABABACACBACAACAB). **B:** Random (frustrated) sequence (ABBBABBABAABBBAAABBABAABABA). The conformation shown is the native (minimum energy) state for each.

MODEL AND METHODS

For the simulations in this work, the model used is a three-dimensional copolymer, which has been described in detail elsewhere.[29,30] The polymers are represented as self-avoiding walks on a simple cubic lattice. Figure 1 shows maximally compact conformations for a 27-unit polymer. We want a potential that can favor compact states, causing the chain to collapse and fold, i.e., a potential that mimics the hydrophobic effect, which is the dominant force for protein folding.[31,32] In our simulations, this effect is modeled by using an attractive potential that favors, to varying extents, the formation of contacts. The potential consists of a contact interaction; monomers that are nearest neighbors on the lattice interact with some energy. Additionally, the potential must encode a unique ground state via the sequence of monomers. We include this effect by making the energy of interactions different, depending on whether the monomers in contact are of the same or different types. The potential energy is given explicitly by

$$E = N_l E_l + N_u E_u$$

where N_l is the number of contacts between monomers of the same type (like contacts) and N_u the number of contacts between monomers of different types (unlike contacts). We are interested in studying simple models, so we use a small number of monomer types; either two or three different monomers, which we label A, B, or C.

Although it is simpler to write the energy function in terms of the E_l, and E_u variables, it is easier to analyze the behavior of the model by considering an equivalent set of parameters, $\overline{E}$ and E_{het}, defined as follows:

$$\overline{E} = \frac{1}{2}(E_l + E_u)$$

$$E_{het} = (E_u - E_l).$$

Figure 2 shows a graphic representation of the connection between the two sets of parameters. $\overline{E}$ represents the overall drive toward forming contacts or compacting the chain. If it is less than zero, contact formation is favored, and the chain collapses. It plays the role of the hydrophobic effect in the present model. E_{het} determines the heterogeneity of the different residues and allows some sequences to encode a unique ground state. In the limit that $E_{het} = 0$, the model becomes a homopolymer. The two energy parameters are each linked to the order parameters (**Q**, **Z**) described above in the following sense. Increasing them increases the drive *along* the corresponding order parameter. Increasing $\overline{E}$ increases the general drive toward compactness, that is, increasing **Z**, the total number of contacts, irrespective of the correctness. Similarly, increasing E_{het} increases the drive toward nativeness, which is measured by **Q**, the number of correct contacts. However, it is important to distinguish setting the value of E_{het} from designing a sequence. For any nonzero value of E_{het} we can design a sequence that folds better than a random sequence, that is, sequences can be designed for any choice of E_{het} (other than it must be nonzero, otherwise we have a homopolymer that is undesignable). E_{het} is not the energy gap often talked about in protein folding theories. This gap depends on E_{het} but also depends on how minimally frustrated the sequence is.

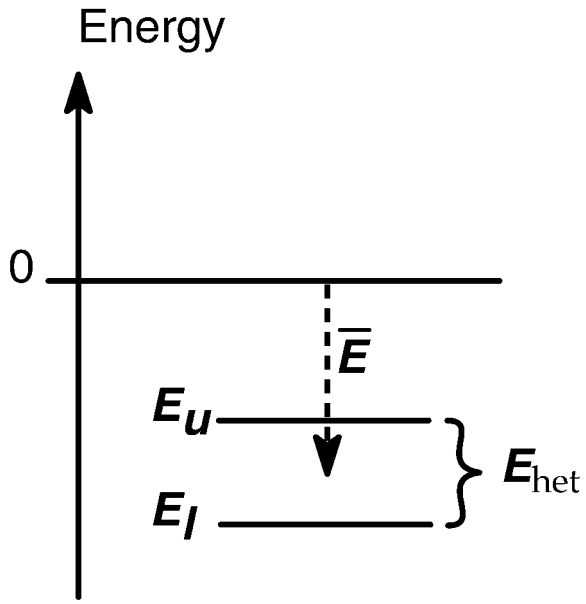

Fig. 2. Energy level diagram. E_l and E_u are the contact energies used in specifying the potentials. $\overline{E}$ represents the average drive for forming contacts or compacting the chain. E_{het} is the splitting in the energy levels and gives the heteropolymer character to the model. $E_l = \overline{E} - E_{het}/2$ and $E_u = \overline{E} + E_{het}/2$.

In this work we are interested on how the model behaves as we change the collapse driving force. We would like some sort of dimensionless parameter, which gives the strength of this collapse force. We define the collapse parameter to be

$$\kappa = \frac{-\overline{E}}{E_{het}}.$$

By adjusting κ we change the strength of the drive toward compactness. We examine two interesting regions of this parameter space: the high hydrophobic limit with $\kappa = 1$ and the low hydrophobic limit with $\kappa = 0$. We use the terms *high* and *low hydrophobicity* to indicate the nonspecific collapse tendency (i.e., collapse into a compact but not necessarily

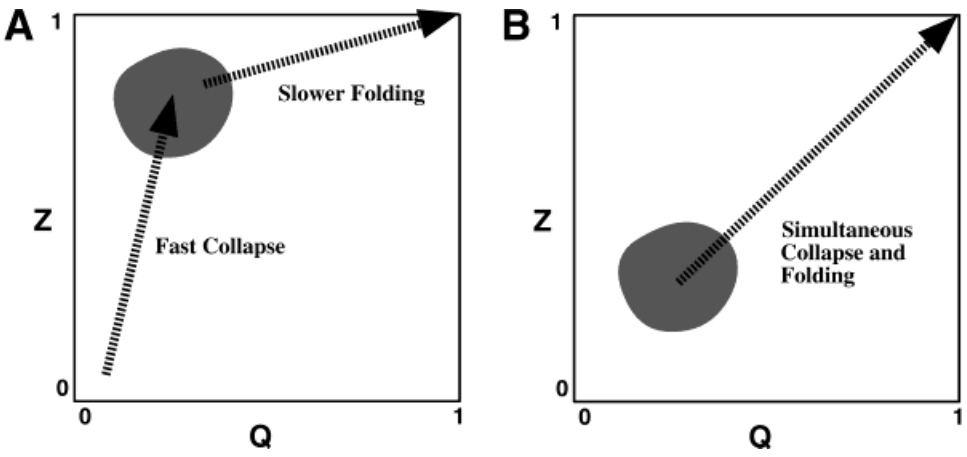

Fig. 3. **A:** Folding scenario for high ($\kappa = 1$) hydrophobicity. **B:** Folding scenario for low ($\kappa = 0$) hydrophobicity.

native structure). By adjusting κ we can modulate this drive toward compactness, which, for real proteins, would be connected to both the composition of amino acids and the makeup of the solvent, that is, one might be able to change κ by changing the number of hydrophobic monomers or by adjusting solvent (by adding ions, or alcohol, etc.). One expects that increasing the strength of the hydrophobic effect would increase the drive toward compactness. Since we keep the composition of monomers fixed in our study, by adjusting κ we are in effect adjusting the solvent properties of our model (which we include in an implicit way because our model does not have any explicit solvent molecules).

At first it may seem odd that the chains still have stable native states with $\kappa = 0$. This is possible for minimally frustrated chains, as the ground state has only contacts between monomers of the same type. So even with $\kappa = 0$, $E_1 < 0$ and the ground state is stable. However, as we will see shortly, this case has very different folding behavior from the high hydrophobic case.

RESULTS

One powerful feature of simple lattice simulations is that one can fully explore the properties of the system as a function of its parameters. It was shown previously[29] collapse occurs either before folding or simultaneously with it depending on the average value of the contact energies, κ. In the regime of large, negative κ there are two distinct events: rapid collapse followed by a slower folding to the native (ground) state. The equilibrium folding and collapse transition temperatures are well separated. As κ approaches zero, folding and collapse occur almost simultaneously.

There has been some confusion and disagreement as to the relation between collapse and folding for proteins. There are many experiments[33,34] suggesting that collapse occurs rapidly and is followed by a slower folding step. However, some recent experiments[35,36] indicate that for some proteins collapse and folding occur together. Various theoretical works have also indicated that both of the different scenarios may occur.[10,14,15,29,37–39] Previous studies of Socci and Onuchic[29] indicate that we can change from one case to the other simply by adjusting one parameter related to hydrophobicity. The different behaviors seen experimentally are probably also due to this effect. To elucidate these different regimes further, we examine two limiting cases of high and low hydrophobicity. For the high hydrophobicity limit, we set the collapse parameter κ to 1. We have done extensive simulations at this parameter value[13,29,30] and have drawn a correspondence between these simulations and small ($\approx$60 amino acids), fast folding proteins.[10,14] To illustrate the low hydrophobicity regime, κ is set equal to zero. Collapse is now driven by the favorable attraction between monomers of the same type alone, with contacts between dissimilar monomers disfavored. This results in far fewer incorrect (nonnative) contacts and more facile reconfigurational diffusion which decreases the folding time. Also compact nonnative states are disfavored energetically so, as the chain collapses, it tends to fall into nativelike states. Figure 3 shows a schematic of the two folding scenarios. The axes show the two-order parameters. **Z** measures the compactness of the chain and is equal to the total number of contacts divided by the maximum number of contacts possible (28). **Q** measures the similarity to the native state and is equal to the fraction of native contacts (contacts that are made in the native state). Both vary between 0 and 1 inclusively. Note, in this paper we are using fractional values for these order parameters. Others (including us[13]) sometimes use integer values for them, in which case they would vary from 0 to the 28.

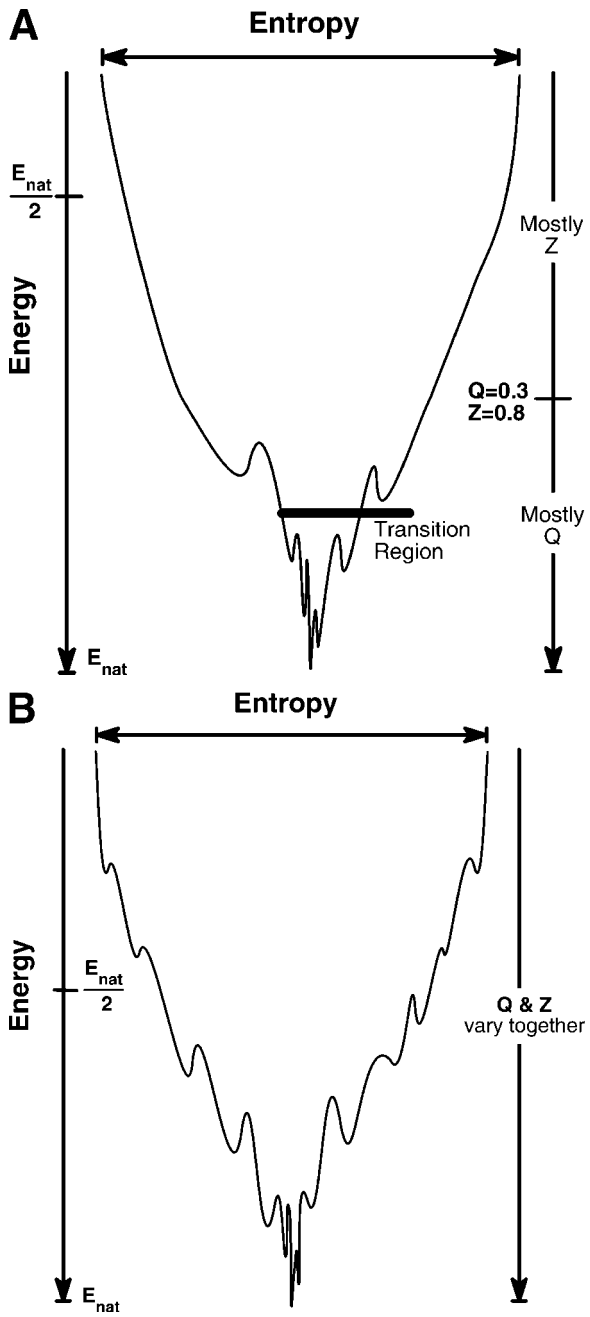

Fig. 4. Schematic folding funnels for the minimally frustrated sequence at high hydrophobicity and low hydrophobicity. The high hydrophobicity case at T_f, where there is an initial rapid collapse to a compact nonnative ensemble followed by slow folding to the native state. The low hydrophobicity case slightly below T_f, where folding and collapse happen simultaneously. The axis on the left of each figure shows roughly the energy (that is, the free energy averaged only over the solvent coordinates of a given configuration) down the funnel, and the axis on the right shows the **Q** and **Z** coordinates. The width of the funnel is proportional to the configurational entropy (the logarithm of the number of conformations).

We point this out to avoid any possible confusion. Another way to view the two folding scenarios is shown in Figure 4. These are schematic views of two different folding funnels corresponding to the high and low hydrophobicity cases of the minimally frustrated sequence. The left-hand figure shows the high hydrophobicity case which is a two-step process. First there is a rapid collapse to a compact, high **Z**, nonnative, low **Q**, intermediate, that is, for the first half of the vertical decent down the funnel, it is the **Z** parameter that is changing. At the molten globule minimum $\mathbf{Z} \approx 0.8$ and $\mathbf{Q} \approx 0.3$. This process resembles a second-order transition and there is no free energy barrier for the transition. This is because the lattice chain is flexible. Stiff chains may exhibit a weak first-order transition and therefore may have a separate thermodynamic barrier. This would be reflected in a narrowing of the funnel in the low **Q**, moderate **Z** region.[40,41] The second stage is the rearrangement of the polymer to the native conformation. Here **Z** is roughly constant and it is now **Q** that is changing. This corresponds to precisely what is shown in the left-hand part of Figure 3. Note that, at the folding temperature, there is transition barrier in the free energy, and this process is first-order–like. In the low hydrophobicity case, the right-hand part of Figure 3, collapse and folding occur simultaneously; both **Q** and **Z** vary by approximately the same amount as we move down the funnel. Consequently, more rapid loss of energy in the high hydrophobicity case cannot compensate for the loss of entropy toward the bottom of the funnel and the result is a barrier in the free energy, which is denoted by the transition region bar. For the low hydrophobicity case, at the temperature for which the figure was drawn (slightly below the folding temperature T_f), the drop in energy is more constant and cancels the loss in entropy so there is no thermodynamic barrier; that is, folding in this case is downhill.

Additionally, for the high hydrophobicity case (fast collapse) we examine two different sequences: a minimally frustrated (good) sequence and a random (poor) sequence. The minimally frustrated sequence has only good contacts (contacts between monomers of the same type) in its folded state; it is completely unfrustrated. Therefore it has the lowest possible energy for its folded state. This was the sequence we used in previous work connecting the model simulations with real proteins. The other sequence was chosen at random. It has a frustrated ground state with a higher energy than the good sequence. This sequence is a nonfolding one with a folding temperature that is lower than its kinetic glass temperature.[30] Consequently, at temperatures low enough for the ground state to be stable, the folding time becomes very long, making this state inaccessible.

Folding Kinetics: Exponential Versus Nonexponential Behavior

Based on a simple dynamical model, it was suggested[1,2] that protein dynamics can be dominated by "hopping" between different low-energy conformations. If the energy landscape is not too rough, this motion can be described as diffusion of a particle over

a free energy surface,[13] which is a function of the folding-order parameter(s) as described above. This description should be a good approximation to the dynamics as long as the energetic roughness is not too large relative to the temperature and stability, or equivalently, as long as folding occurs through a large number of different pathways,[42,43] that is, only short-lived traps exist, and as the protein escapes from these traps it can go to a large number of different configurations. This description holds as long as the temperature remains above the glass transition temperature (T_g) as discussed by Bryngelson and Wolynes.[1] For temperatures below T_g, the proteins have trouble escaping from a few specific long-lived low-energy traps and, when that occurs, it will pass through a small number of configurations that are connected by "untypically" small barriers. Because of that, the many-pathways description breaks down, and most of the allowed paths do not lead the protein toward the native state. This transition from a many paths funnellike situation toward a few largely nonnative states with specific paths is the kinetic consequence of the glass transition. If trapped on the "wrong pathway," the protein must nearly unfold before it can fold successfully, leading to extremely long folding times.

The ruggedness of the energy landscape of a heteropolymer is reflected in its glassy behavior. The qualitative hallmarks of the behavior of glasses as seen in the laboratory study of bulk systems are worthy of note here: first, the slowing down of configurational dynamics with decreasing temperature (as reflected in the viscosity), and second, the complex nonexponential time course of structural relaxation of highly viscous liquids. The glass transition within individual folded proteins has been observed experimentally.[44–46] Also, the glass transition in bulk protein materials has been directly observed.[47,48] Indeed those studies confirm T_f exceeds T_g. One would expect that the bulk glass transition occurs at higher temperatures than for an isolated globule, strongly supporting the minimal frustration notion, which holds that T_f is larger than T_g also for the individual protein molecules. The slowing of all configurational changes as T_g approaches is the cause of the nonmonotonic course of the folding time with temperature decrease, well documented in numerical simulations.[6,13,29,30,49,50] An important point that should be considered when analyzing these results is that the effective interaction parameters between residues are themselves temperature-dependent because of the hydrophobic effect. This is also a source for the slowing of the folding time with decreasing temperature, that qualitatively may be included in the lattice simulations.[9,17,23] Nonexponential kinetic behavior is also often seen. Here we explore the relation between the glass transition and this nonexponential behavior.

In thinking about glassy phenomena, several ways of defining characteristic temperatures need to be considered. For macroscopic laboratory systems, a kinetically defined glass transition is the least controversial. This kinetic T_g is defined as the temperature at which the relevant correlation time reaches some large (but somewhat arbitrary) value set by observation time scales. For laboratory glasses this temperature, T_g, empirically correlates with an ideal glass transition temperature T_k, where thermodynamically the configurational entropy per particle would vanish, if equilibrium could be maintained. Likewise for heteropolymer simulations, a kinetic T_g can be defined once an observation time scale is decided upon. Even though the precise value of T_g depends on the moves used and on the choice of observation time, the kinetic T_g only depends weakly on these choices (see ref. 30 for dependence on the observation time) and is similar to the T_g (also called T_k) value given by the random energy model proposed of Bryngelson and Wolynes.[1] This ideal glass transition temperature T_g is related directly to the energy fluctuations in the collapsed state and its entropy. In ref. 13 this random energy model estimate,

$$T_g^{\mathrm{REM}} = \frac{\Delta E}{\sqrt{2S_L}}$$

was made for the minimally frustrated sequence using values for the energy fluctuations and configurational entropy measured for the collapsed states. In the same paper, it is shown that this estimate also agrees with the critical temperature where replica symmetry breaking sets in. This connection is fully explored throughout this subsection of the article. (As discussed in ref. 10, the replica symmetry-breaking temperature, T_g (**Q**,**Z**), can be estimated by computing the number of thermally accessible microstates for a fixed choice for **Q** and **Z**. This can be done with little ambiguity even for finite system (see the subsection Kinetic View: Y[**Q**,**Z**] and Folding Trajectories below for details). A more precisely defined T_g(**Q**,**Z**) would use the configurational entropy of basins of state. For a finite system this has a little arbitrariness involving the choice of cutoff similarity measures for basins. The close comparison between the ideal thermodynamic and the kinetic glass transitions provides strong support of the connection between thermodynamic trapping and the kinetic transition from many to few pathways described above.

Before we present our simulation results and their interpretation, using modern theory, we describe some recent advances that go beyond the random energy model (REM) formalism. Plotkin and colleagues[51] have developed the correlated landscape approach for the random heteropolymer. A main conclusion of this work it that, even for the worst case of pair interactions alone, the glass transition temperature is actually very little changed by the correlations, thus justifying the use of simpler theo-

ries. There is, however, one significant modification, equally important as predicting the appropriate free energy profile, there is the need of understanding the configurational diffusion coefficient that reflects the sampling of transient trap structures. Within the REM approximation this reflects the full energetic ruggedness of the landscape, while in the generalized random energy model (GREM) the barrier is much reduced.[13] In a correlated landscape, the GREM illustrates that there is another characteristic temperature, which is called T_A. Below T_A, activated dynamics of escape from traps becomes important and has been implicated in theories of structural glasses. Takada and Wolynes,[52] using a replica technique, estimated the barriers between the onset of activated dynamics (T_A) and the glass transition at T_k. Plotkin and coworkers[53] showed the GREM also exhibits the two transitions at T_A and T_k and that the correlations diminish significantly the barrier heights. T_A is relatively controversial. Both analytical approaches mentioned, based on correlated landscapes and replica methods, give values for lattice models of the order $1.4T_k$. A mode coupling calculation of Roan and Shakhnovich[54] led them to suggest T_A does not exist, while another calculation by Thirumalai and coworkers[55] does provide a value for it that depends on chain length. Dawson and coworkers[56] have carried out a numerical treatment of collapse dynamics which shows a glass transition. Takada and colleagues[57] recently developed a mode-coupling calculation compatible with the replica theory results.

We now quantify the glass transition using the simulations of finite proteinlike sequences described in this section. We must bear in mind, however, that it is difficult to define crisply any phase transition for a small finite system. This is more true for the dynamical transition than for the equilibrium transition. But the two rough criteria already introduced: the kinetic one signaling a transition at "laboratory" time scales defined by the folding time itself, and a simplified replica symmetry-breaking calculation, connected with energy traps, can be used in understanding the simulation data (see the subsection Kinetic View ... below). The latter signals the transition to a "few state behavior;" i.e., an entropy crisis, and corresponds to what is calculated via replica calculations and can be taken to represent the transition to specific intermediates. These results are then discussed and quantitatively analyzed in the light of the analytical theories.

To obtain a better understanding of the kinetic consequences of the glassy effects, we use our simple lattice model to probe the relation between dynamical slowing down and nonexponential kinetic behavior. We measured the fraction of the population that remains *unfolded* as a function of time. This is a quantity that can be inferred from folding experiments. The value of some probe (typically a spectral one), which has two different values in the folded and unfolded state, is monitored as a function of time. The percentage of the population that has not yet folded can be determined and can be plotted on a log-log plot.[58–60] This sort of plot introduced by Frauenfelder and coworkers[59] for protein dynamics is very convenient because a simple exponential process has a very characteristic and easily identifiable signature and can be distinguished from nonexponential behavior while power-law dependencies, reflecting a broad distribution of lifetimes, give rise to linear curves. We perform these studies for several sequences with different folding characteristics. Figure 5 shows these traces for a minimally frustrated, three-letter code sequence with energy parameters that favor collapse (the high hydrophobicity case). This is the system we have previously discussed as being analogous to a small, highly helical, fast folding protein sufficiently hydrophobic to ensure fast collapse.[10] Figure 6 shows similar plots for the same sequence but with the low hydrophobicity energy parameters, which no longer favor rapid collapse, but rather simultaneous collapse and folding. Here the correspondence with a laboratory protein is no longer quantitative since additional entropy is liberated by the expansion of the molten globule. Finally, Figure 7 shows the results for a random (frustrated, but with a single ground state) sequence at high hydrophobicity.

Before analyzing these results, we introduce the notion of "effective Levinthal time." In a simple analysis Levinthal[61,62] supposed the protein would have to search all its configurational states in order to find the lowest energy one. If such a situation were correct, it would take a prohibitive amount of time for proteins to fold. This gave rise to the Levinthal paradox. The solution of this paradox is found in the fact that this would be true only if the landscape had absolutely no bias favoring the native (folded) state. This is clearly not the case for good sequences satisfying the principle of minimal frustration. It is also important to keep in mind, that for finite systems, even when there is no bias, this Levinthal time only represents an average time to reach a chosen configuration. Depending on the dynamical connectivity, different states may have different accessibility. This was discussed previously[2] and later clearly shown by Betancourt and Onuchic[63] for folding simulations in 2D-lattices. Additionally, it was shown even for a homopolymer the random search time is not computed correctly by the Levinthal argument and that the Levinthal time does not give the folding time even in the absence of a bias.[6] Still, the Levinthal time concept is a benchmark for understanding the typical folding time for a random (frustrated) sequence. For the kind of sequences we have been studying, however, the standard definition of the Levinthal time is not appropriate. Even for random sequences, our choice of potentials favors

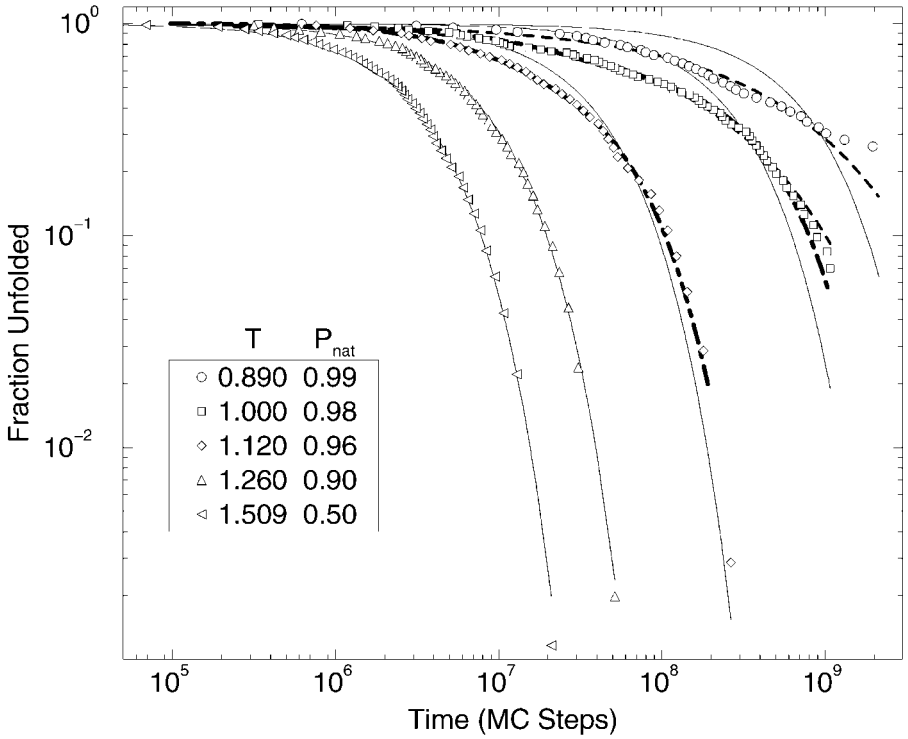

Fig. 5. Fraction of population left unfolded versus time; plotted on a log-log scale. The sequence is the minimally frustrated three-monomer–type code described in detail in ref. 10. The energy parameters correspond to the high hydrophobicity regime favoring collapse. The plots were generated by running many folding simulations (typically from several hundred to as many as 1000) at a given temperature and recording the folding time. Then at each time the fraction of the number of runs that have not yet folded is plotted. Note that, once a chain folds, the simulation is stopped, so we do not allow it to unfold, i.e., there are no back reactions from folded to unfolded (as would certainly occur in a real experiment on an ensemble of molecules; however, single molecule experiments can monitor this survival fraction directly). The symbols show these data for several temperatures. The highest temperature correspond to the folding temperature (the temperature at which the fractional equilibrium native population equals one-half, $P_{nat} = 0.5$) for this sequence ($T_f = 1.51$). The kinetic glass temperature is approximately 0.9 ($P_{nat} \approx 0.99$). The thin solid lines are single exponential fits to the data: e^{-t/τ_f}. The dot-dashed lines at $T = 1.00$ ($P_{nat} = 0.98$) and $T = 1.12$ ($P_{nat} = 0.96$) are double exponential fits, and the dashed lines at $T = 1.00$ and $T = 0.89$ are stretched exponential fits: $e^{-(t/\tau_f)^{\beta}}$ with $\beta \approx 0.78$ at the $T = 1.00$ and $\beta \approx 0.45$ at the lowest temperature. The time is given in Monte Carlo steps.

collapse, and the folded states are always chosen as fully collapsed ground states of the system. Thus, it is convenient to call the "effective Levinthal" time for a given choice of potentials, the mean folding time of a random sequence at its folding temperature T_f, the temperature that the folded structure becomes stable. At T_f, good folding sequences fold on times much shorter than the effective Levinthal time which is characteristic of random sequences. The effective Levinthal time is more complex than a simple product of the number of states that need to be visited times some simple reconfigurational time step, τ_0. By assuming a discrete dynamics, the usual argument ignores the existence of a full spectrum of Rouse modes.[64] Most important, however, it neglects the effect of steric constraints on the moves. For the collapsed states, most of the move trials are not accepted. No complete analytical theory that addresses how these factors will change or renormalize the fundamental time step has yet been developed.

As discussed above, Figure 5 shows the nonfolded fraction as a function of time for the minimally frustrated sequence in the high hydrophobicity case. Data for a range of temperatures is plotted, from the folding temperature T_f (1.51) down to below the kinetic glass temperature T_g (0.96) for this sequence. For this sequence collapse occurs early in the folding event, and most of the time is spent searching collapsed configurations. At the folding temperature, the kinetics of the system is clearly single exponential, indicating that folding is a two-state process. Kinetically, two-state behavior is seen in the experiments on small fast folding proteins. In these simulations the denatured state itself is not expanded but can be reached by a downhill flow from the extended chain.[33,65–68] No long-lived traps are populated during the folding process. As discussed by us earlier,[13] only the average effect of these nonspecific short-lived traps is important in determining the folding kinetics and it can be included via an effective configurational diffusion coefficient. Recently, several experiments have been directly measuring this diffusion coefficient,[69,70] providing strong support for this view of the folding mechanism. As the tempera-

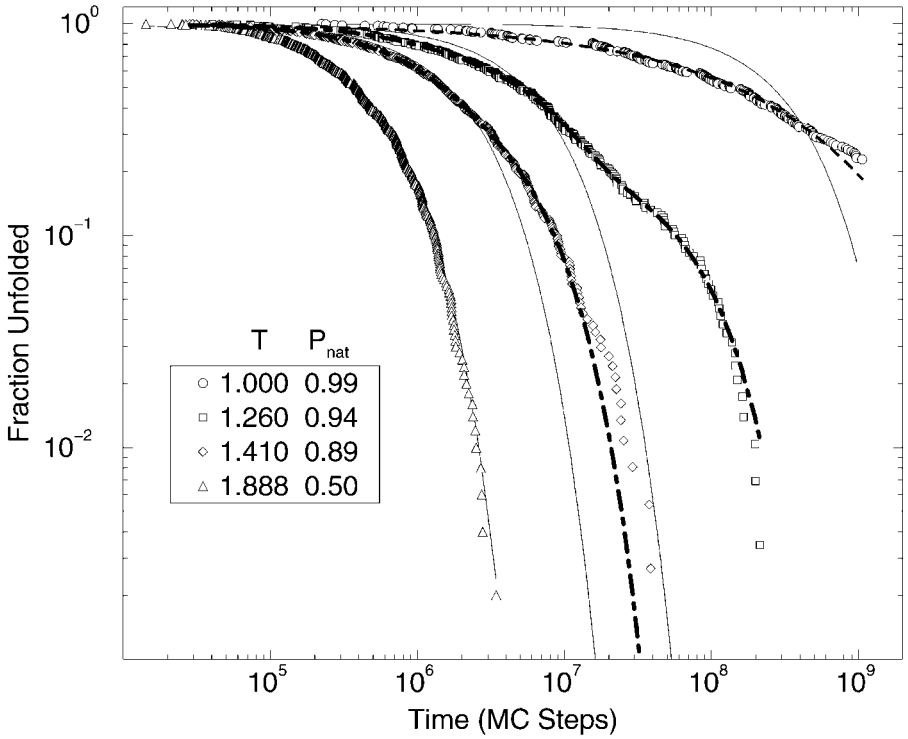

Fig. 6. Fraction unfolded versus time, for the minimally frustrated sequence at low hydrophobicity energy parameters. The folding temperature is $T_f = 1.89$, and it is only at this temperature that the kinetics are single exponential. The thin solid lines are the single exponential fits. The dot-dashed lines are multiexponential fits, double for $T = 1.41$ ($P_{nat} = 0.89$) and triple for $T = 1.26$ ($P_{nat} = 0.94$). The dashed line at $T = 1.00$ is a stretched exponential fit with $\beta \approx 0.4$.

ture is lowered the kinetic behavior becomes more complex. As the temperatures approach T_g (around 1.1), the time course becomes multiexponential. Initially upon cooling below T_f the kinetics can be fit well with two exponentials, indicating seemingly that only a few long lived traps start to appear. Complex kinetic behavior has also been seen previously in folding experiments on real proteins[71,72] and was also fitted with a multiexponential form. Below $T = 1.1$, the range of lifetimes of kinetic traps rapidly increases to the point that when the kinetic glass temperature ($T_g = 0.96$) is reached, the curve can be approximated with either an explicit multiexponential or more economically as a stretched exponential $e^{-(\tau/\tau_0)^\beta}$ function. For large β ($\approx$0.7) two or three exponentials are sufficient; however, for smaller β's ($\approx$0.45), it would take 5 to 6 exponentials to fit the stretched form over 3 decades. When the kinetic glass temperature is reached, the folding mechanism switches from the situation where many equivalent pathways involving many states of similar lifetimes are averaged over to the limit of a "few" kinetically inequivalent pathways. (Notice that "few" still is a reasonably large number of pathways, but this number is much smaller than the exponentially large number of states.) Also, since protein systems are finite, as the temperature continues to be reduced below T_g, the long time behavior is dominated by a single exponential again because it will get controlled by the deepest trap. The folding time, in any simulation, now depends on the specific pathway on which the protein was initially trapped. Recall that if the protein gets into a pathway that is not connected to the native state through the compact manifold, it may have to almost unfold before trying to fold again. This not only slows substantially the folding process but results in strongly nonexponential behavior. Different traps, that is, different pathways, control folding for different folding events.[42,43] This is what one would expect near a glass transition. Thus there is a connection between the theoretically predicted ideal glass temperature and the kinetic one. As this temperature is reached, the kinetics is controlled by a number of low energy long-lived traps. This is exactly what we expected from our unidimensional (single-order parameter) funnel described in ref. 10.

Figure 6 shows a similar plot for the same sequence but with the low hydrophobicity energy parameters, which no longer favors initial rapid collapse, but rather simultaneous collapse and folding. The equilibrium denatured state is expanded. The temperatures in the plot again range from T_f down to approximately T_g. However, a seemingly different kinetic behavior emerges. At T_f the system only "collapses" when it folds correctly, and folding is a

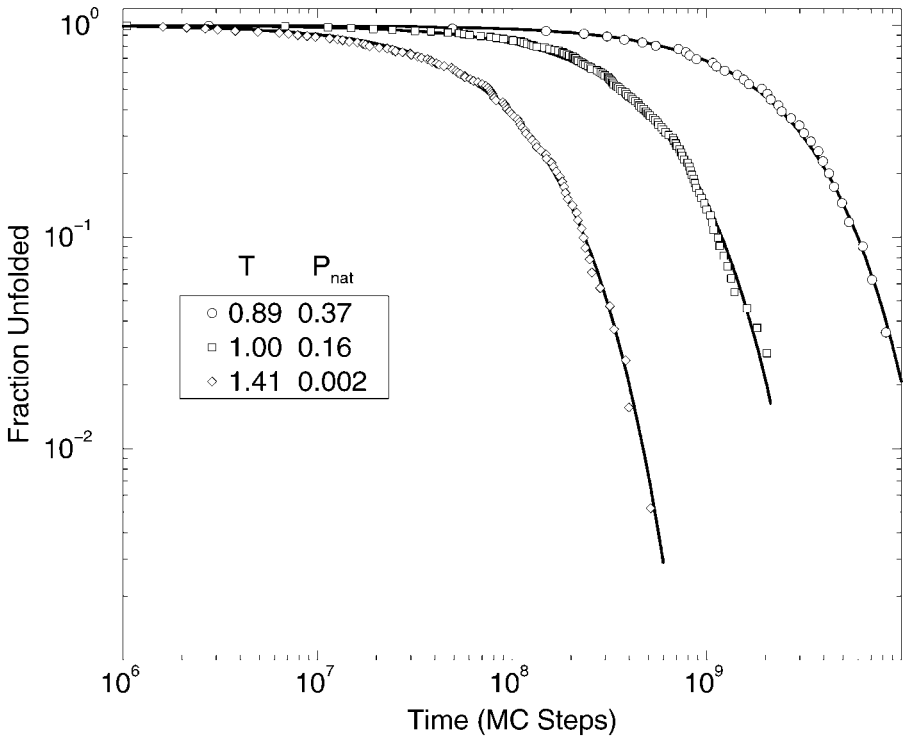

Fig. 7. Fraction unfolded versus time for the random (frustrated) sequence at high hydrophobicity. The folding temperature ($T_f = 0.83$) is quite low, below the kinetic glass temperature ($T_g = 1.0$, $P_{nat} = 0.16$). So all the temperatures shown here are above T_f. The kinetics are single exponential at all temperatures shown.

two-state process with simple exponential kinetics without the initial collapse. The collapse temperature for low **Q** structures is lower than T_f. This is analogous to a substance that sublimes rather than melts before boiling. However, as the temperature begins to approach T_g, the sequence may collapse into compact structures other than the folded one. Naturally, at T_g the system gets trapped in several nonspecific collapsed structures, but multiexponential behavior occurs at higher temperatures too. This is because, as we shall see in the multidimensional funnel, T_g depends on **Q** and **Z**. These collapsed structures are rarely visited at temperatures around T_f. Thus, for the range of temperatures between T_f and T_g, more and more collapsed and nativelike structures are visited as the temperature falls. Such states are responsible for the creation of specific traps on the route toward folding, destroying the kinetic equivalence of the pathways. This leads to nonexponential behavior well above the glass transition temperature defined kinetically by the slowest speed because now we are monitoring the glass transition on the fast folding scale. Here the folding times in Figure 6 are still relatively fast at the temperature where the system becomes multiexponential. Again, if specific traps control the dynamics, the configurational diffusion picture described in the beginning of this section becomes invalid and cannot be used quantitatively.

A comparison between the folding mechanism for the high and low hydrophobicity good sequences is appropriate at this moment. Even though we often refer to a single T_g (the one for $\mathbf{Q} \approx 0$, **Z** highly collapsed), it is natural to expect that trapping may start to occur at different temperatures for different values of the order parameters. This can be understood in the following way. As the protein evolves towards the folded state, its kinetics may not be dominated by long-lived traps during the early events but trapping may occur later on when the number of microstates available is severely reduced. For example, a minimally frustrated sequence at T_f does not get trapped at the molten globule state, but it may find some traps close to folding minimum. The number of traps in this situation, however, is small enough, and does not cause major problems for folding. On the other hand, a random (frustrated) sequence may already hit long-lived traps at the molten globule state around T_f. Since the number of traps is now "exponentially" large, folding becomes problematic.

Trapping is specially sensitive to the degree of collapse **Z**, that is, more collapsed configurations start to get trapped at higher temperatures. This dependence of the glassy behavior on **Z** hardly changes things for the one-dimensional case since many accidentally stable contacts can occur in a collapsed state for the high hydrophobicity sequence.

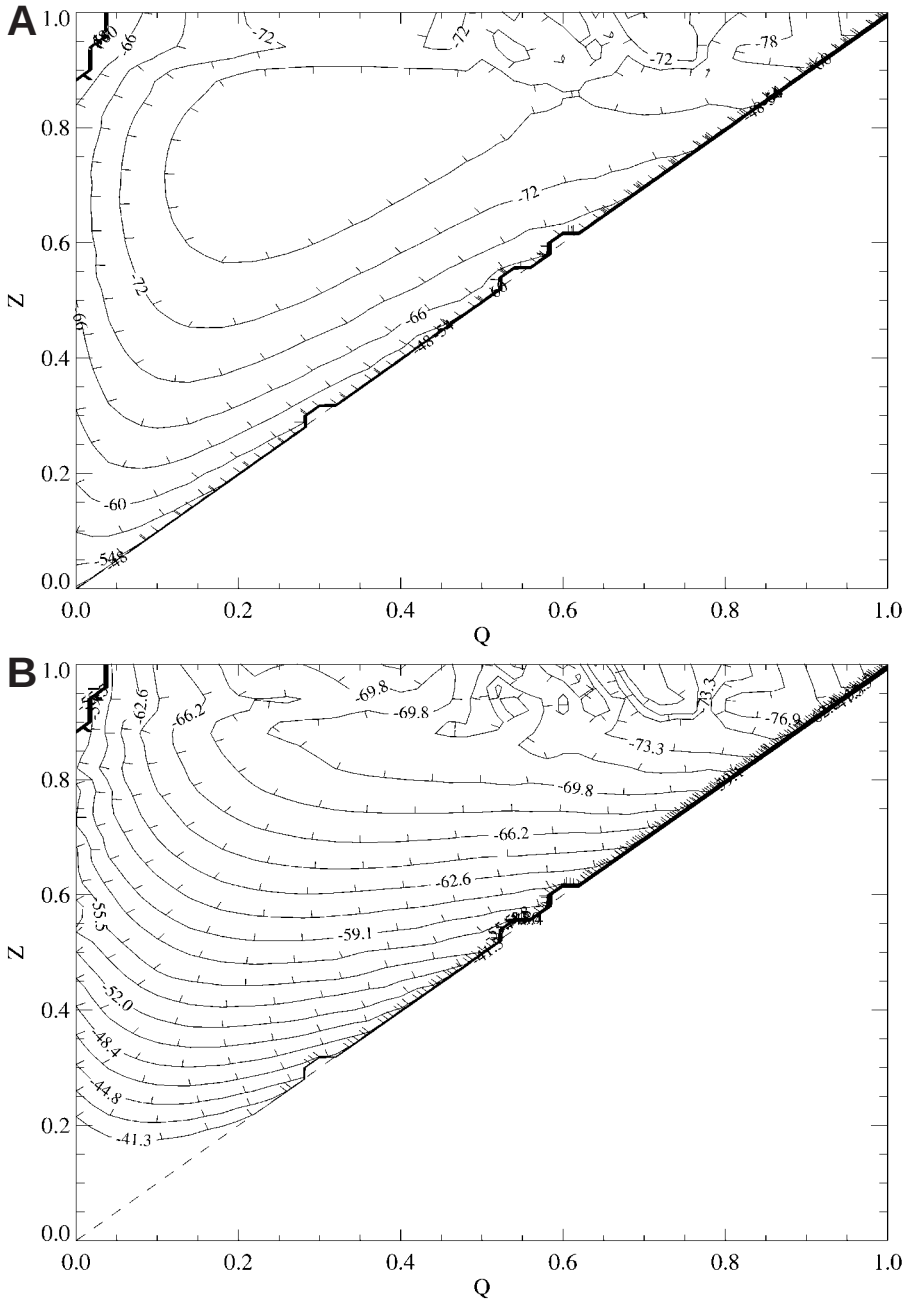

Fig. 8. Free energy as a function of **Q** and **Z**. Good sequence at high hydrophobicity. **A:** Folding temperature. **B:** Low temperature. The contour lines are $2k_BT$ apart. The tick marks indicate the "downhill" direction; i.e., they are in the direction of the negative gradient.

Since collapse occurs early during the folding event, most of the folding trajectory occurs for high values of **Z** and the important dynamics is only one-dimensional. For this reason, the value of T_g almost does not vary, that is, the molten globule T_g is very similar to the one for larger values of **Q**. The situation completely changes for the low hydrophobicity sequence. In this case, **Q** and **Z** vary simultaneously during the folding event. Since the molten globule state is far from being totally collapsed, regions in the folding trajectory with larger **Q** values have a T_g larger than for the configurational space relevant for the molten globule one. The kinetic consequence of these T_g variations is shown in Figure 6. There are temperatures that the protein is not glassy while at the molten globule minimum, but trapping occurs at larger values of **Q** as it is on its way towards the native state. To fully explore and quantify this

dependence of T_g on the order parameters, the replica symmetry breaking temperature, T_g (**Q**,**Z**), is estimated in the subsection Kinetic Views . . . by computing the number of thermally accessible microstates for a fixed choice for **Q** and **Z**.[10] This study allows us to compare these "thermodynamic" glass temperatures to the kinetic simulations presented in Figure 6. The local glass transition temperature depends on how the entropy and ruggedness change with the order parameters. For the lattice models, the entropy decreases faster than the ruggedness. Presumably this is because one can generate rather large energy fluctuations by packing together incorrectly structures that are otherwise properly ordered. It is important to recognize that since entropy has greatly decreased by this point, landscape theory[2] indicates that this glass transition is kinetically weaker involving a much smaller slowing down than if it were to occur at $\mathbf{Q} = 0$.

It was not feasible to characterize all possible behaviors for poor (frustrated) folding sequences. We now explore one of these possibilities. Figure 7 shows the unfolded fraction as a function of time. The energy parameters for this sequence are the ones for the high hydrophobicity case. For this sequence the kinetics remains single exponential even below the ideal glass temperature. Note that the folding temperature ($T_f = 0.83$) for this sequence is actually *below* the glass temperature ($T_g = 1$). Therefore at T_g there is still a barrier; that is, folding is still an uphill process, unlike in the previous cases where folding was downhill by the time the temperature reached T_g. So the folded state of this sequence is still not thermodynamically stable even at this low temperature. T_f being lower than T_g is a finite size effect (infinite systems can not have transition temperatures below T_g[73] since in an infinite system "one" (polynomial number) state will dominate below T_g). This means that the lowest energy state is only stable for temperatures below the glass temperature. Similar to the behavior of the minimally frustrated sequence, as the temperatures approaches T_g, trapping on nonspecific collapsed states occur. However, instead of generating inequivalent pathways, the folding time is now so long (on the order of the effective Levinthal time), that the residence time in these traps is still very short compared to the folding time, that is, the protein "averages over" these deep traps before it folds, and folding remains exponential in time because of that. Or equivalently, since the native state is the "deepest trap," it determines the folding time, again leading to the exponential behavior as a finite size effect.

In subsequent sections we examine both the free energy and the replica breaking parameter ($Y = \Sigma_i P_i^2$), which provides insight into the different behaviors of the various cases.

Thermodynamic Properties: Free Energy

In nature, does collapse occur before folding or do folding and collapse occur simultaneously? There is experimental evidence for both scenarios in different systems. In the simple lattice system we study, it is possible to interpolate between the two regimes by adjusting the average energy of the interactions ($\overline{E}$). We can immediately appreciate the origin of the two behaviors by examining the free energy of the system as a function of the two order parameters: **Q** which measures the number of correct contacts, or how "folded" the conformation is, and **Z** which measures the total number of contacts, or how collapsed the conformation is. Figure 8 shows the free energy surfaces for our good sequence for the high hydrophobicity or strongly collapsing condition. At the folding temperature there are two minima in the free energy profile: one is the folded state (at **Q** and $\mathbf{Z} = 1.0$) and the other is an ensemble of compact states with roughly 75% of the total number of contacts but with only 30% of the correct contacts made at any time. The behavior of this system is kinetically two-state–like and folding here consists of diffusion about these compact states which eventually surmounts the barrier separating them from the folded state. As we showed previously[13] the rate-limiting step is crossing the thermodynamic barrier. Recent all atom simulations for segment B1 of *streptococcal* protein G provides evidence that its folding falls in this regime.[37] At low temperatures, the separated collapsed minimum and the barrier disappear and the free-energy surface is downhill (a type 0 scenario in the language of ref. 9). The folding time is determined by the diffusion constant along this surface, but at this temperature the diffusion constant is quite small and the folding time is large even though the free-energy surface is downhill.

If we now look at the same sequence but under the low hydrophobicity weakly collapsing parameters the free energy surface is very different (Fig. 9). At the folding temperature it still has two minima, but now the non-folded minimum is not collapsed. The value of **Q** is still about 0.3 but the value of **Z** is reduced to roughly 0.45. Not only are the conformations in this minimum less collapsed but since $\mathbf{Q} \approx \mathbf{Z}$ the systems tends to only make the correct, native contacts. Collapse and folding occur simultaneously. As we will see in a subsequent section the folding trajectories tend to follow the line of $\mathbf{Q} = \mathbf{Z}$ (see Fig. 3). Although there is a barrier, it is rather small, and as the temperature is lowered the barrier disappears. In this case, the downhill region occurs at temperatures high enough that the diffusion constant is still large so folding is still very fast. As opposed to the high hydrophobicity case where the downhill region occurs only at fairly low temperatures with a small diffusion constant. Also, the majority of the maximally compact states ($\mathbf{Z} = 1$) are

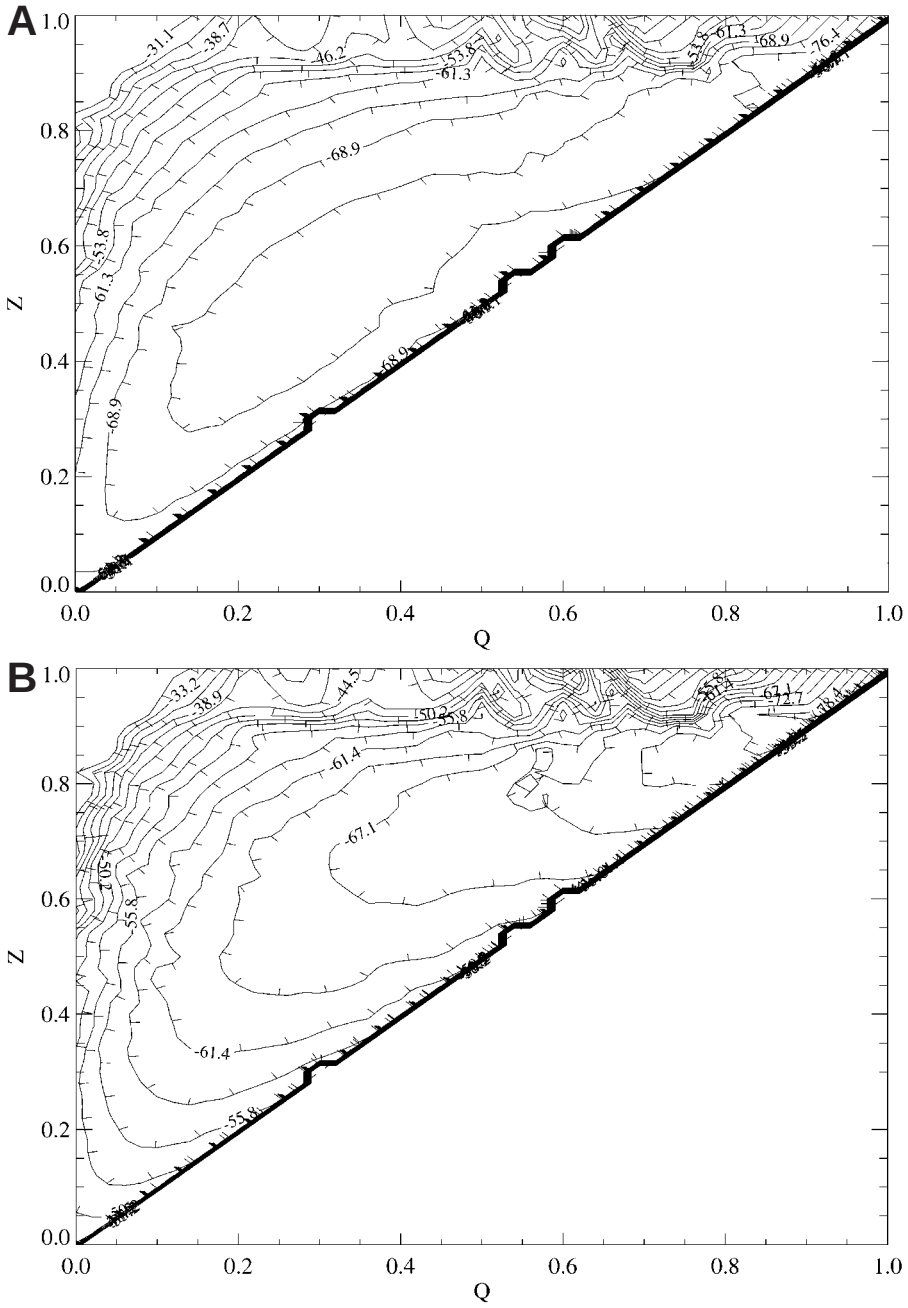

Fig. 9. Free energy as a function of **Q** and **Z**. Good sequence low hydrophobicity. **A:** Folding temperature. **B:** Intermediate temperature between T_f and T_g; at this temperature the kinetics are no longer single exponential. **C:** Approximately kinetic glass temperature. Note that at the folding temperature there is a small barrier, while at the lower temperatures the barrier is no longer there. Again the contour lines are $2k_BT$ apart.

"uphill" and consequently do not form traps in this case. Recent all atom simulations for the folding of a three-helix bundle protein indicate that its folding is probably in an intermediate regime between the high and low hydrophobicity conditions.[38] Later in this article, we see that the traps in this case are so called on-flow (as opposed to off-flow in the previous case).

The last case we examine is for the frustrated (random) sequence at high hydrophobicity. At high temperatures, the free energy surface (Fig. 10) is very similar to the one for the good sequence at high hydrophobicity (Fig. 8). There are two thermodynamic minima: one is the native state and the other an ensemble of nonnative (low **Q**) collapsed (high **Z**) structures. The barrier for this sequence is much

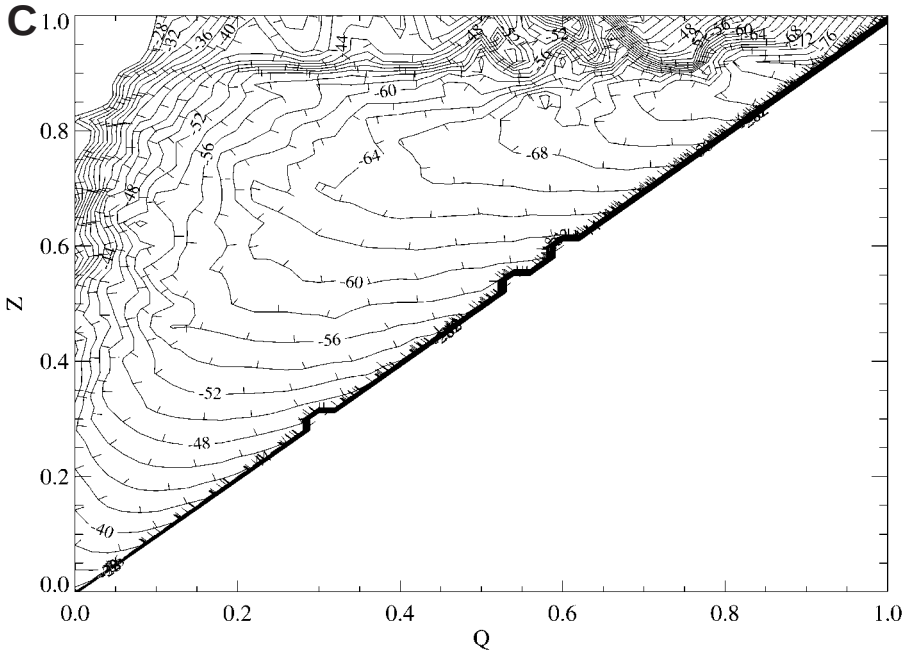

Fig. 9. (Continued.)

higher and consequently the folding time is much longer. At low temperatures (below T_g) the behavior is now different from the good sequence. There are still two well-defined minima and a sizable barrier between them. This partially explains the folding time course behavior we saw previously (Fig. 7). The system is still exponential even at low temperatures, as there is still a barrier to cross, while the folding time is nevertheless dominated by diffusion about the collapsed minimum.

In fact, if we compare the free energy plots with the folding time plots, we see the following general pattern. If there is a reasonable barrier separating two well defined thermodynamic minima then the folding time course is exponential. When the barrier disappears and folding becomes a downhill process then the time course becomes nonexponential. Depending on the size of the barrier or the diffusion coefficient, the folding time may be fast or slow. This can lead to the counter intuitive cases of very fast folding with nonexponential kinetics and in another case very slow, exponential folding. Of course there is also the usual fast exponential case and the slow nonexponential one. What is particularly satisfying is that this wide range of disparate behaviors can be seen as a result of simply adjusting two parameters: the average drive to compactness and how minimally frustrated a sequence is.

Kinetic View: Y[Q,Z] and Folding Trajectories

In the lattice model, as in real proteins, there is a wide range of kinetic behavior for folding. One aspect involves the various time scales of collapse and folding. We saw in the previous section that by modifying the collapse parameter (κ) we could change the thermodynamic free energy surface from one with rapid collapse before folding to one with concerted collapse and folding. Another aspect involves the detailed time course of folding, that is, simple exponential versus nonexponential. We now explore this aspect in greater detail and look quantitatively at what factors determine this behavior. One useful piece of information is the Y[**Q**,**Z**] plot which is a measure of the number of thermally accessible states. It is defined as follows:

$$Y[\mathbf{Q},\mathbf{Z}] = \sum_{i} P_{i,}{}^{2}\,[\mathbf{Q},\mathbf{Z}]$$

where $P_i[\mathbf{Q},\mathbf{Z}]$ is the Boltzmann probability of microstate (single configuration) i with a given value of the order parameters **Q** and **Z**. Y is equal to the thermal average of the Boltzmann probability:

$$Y[\mathbf{Q},\mathbf{Z}] = \langle P[\mathbf{Q},\mathbf{Z}]\rangle = \frac{1}{N_{\text{therm}}}$$

and is also equal to the inverse number of thermally accessible states denoted N_{therm}. In previous works[10,13] we examined Y only as a function of **Q** and were able to see the appearance of the glass transition and locate the "local glass transition" that occurred for high values of **Q** at T_f. At a fixed temperature we

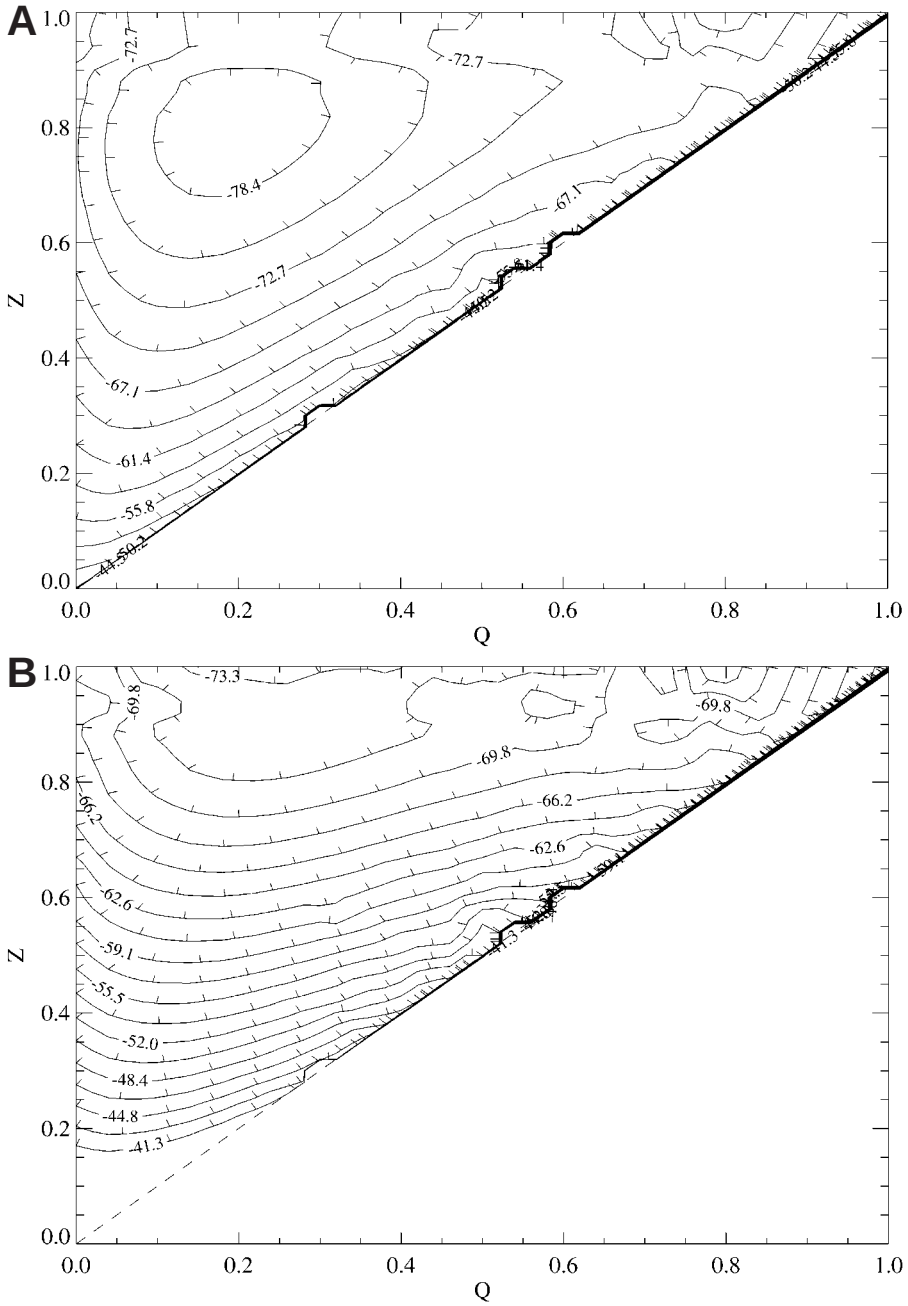

Fig. 10. Free energy as a function of **Q** and **Z** for the bad sequence at high hydrophobicity. **A:** High temperature, where the folding time is the fastest; it is well above the folding temperature. **B:** Kinetic glass temperature, which is slightly above the folding temperature for this sequence. In both cases there are two minima with a barrier between them.

could calculate the **Q** value, $\mathbf{Q}_g$, at which the number of states and Y approached 1. In general we expect glassy behavior when the number of thermally accessible states is small. Just as a two-dimensional plot of the free energy along two order parameters **Q** and **Z** reveals greater detail about the properties of the system, the two-dimensional version of Y is also much more informative.

Figure 11 shows the Y[**Q**,**Z**] plot for the minimally frustrated sequence at the high hydrophobicity value of κ. Both high (T_f) and low (T_g) temperatures are plotted. As Y approaches 1 the contour lines are roughly parallel to the **Q** axis, indicating it is actually the **Z** order parameter that mostly controls the glassiness in the system. By examining the Y plots alone, it is not immediately apparent why the two

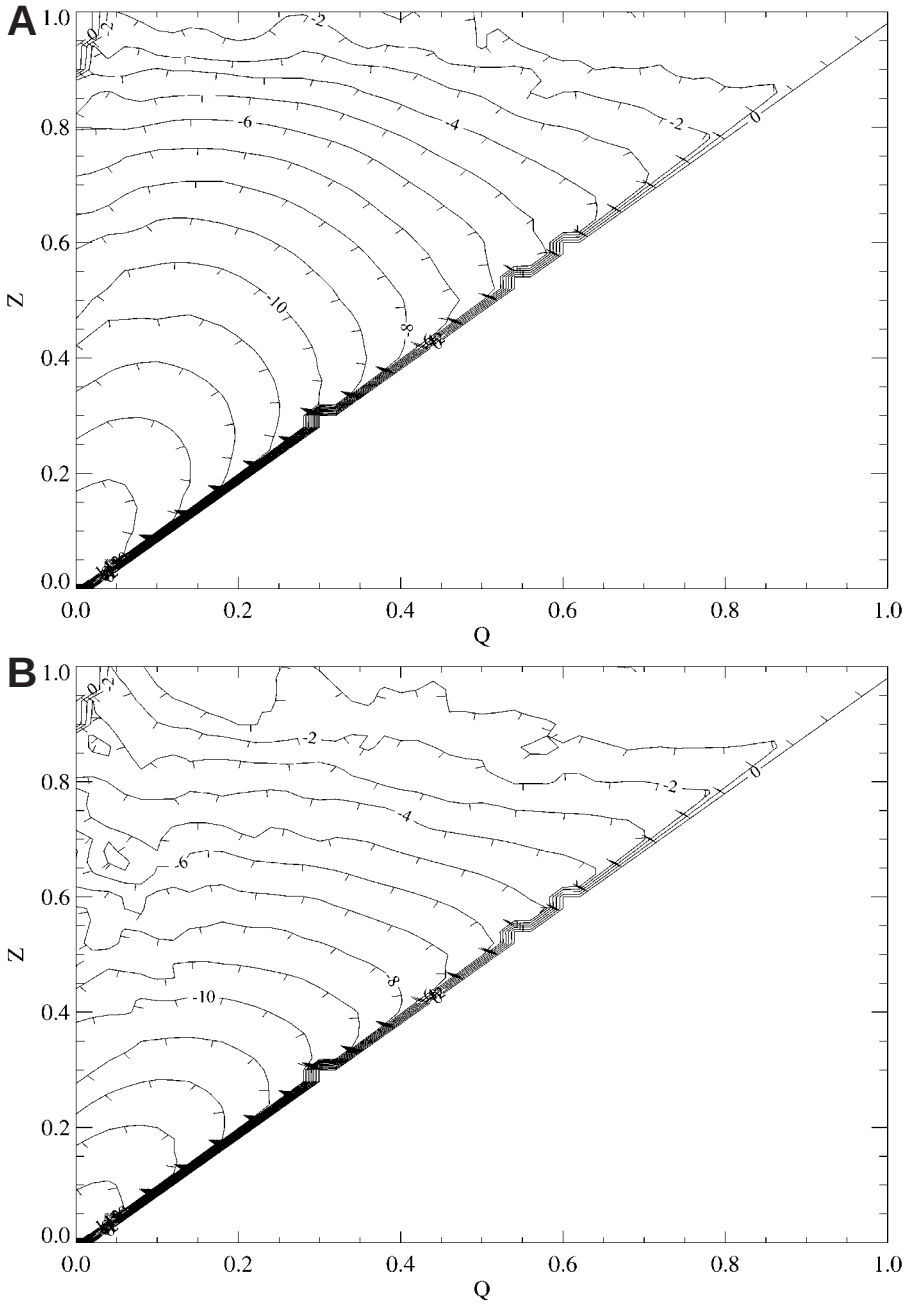

Fig. 11. Replica symmetry breaking parameter Y[**Q**,**Z**] = $\Sigma_i P_i^2$. Minimally frustrated sequence at high hydrophobicity. **A:** Folding temperature. **B:** Temperature below T_g (right). The contour levels are the base 10 logarithm of Y, so a value of -2 indicates $Y = 10^{-2}$.

cases have different kinetic behavior. The key is to consider both the free-energy plots superimposed on the Y[**Q**,**Z**] plots. Shortly we show these plots from our model simulations but first we examine a schematic version of them.

Figure 12 shows a rough plot of the free-energy surface and one line from the Y[**Q**,**Z**] plot; the line $Y = 0.1$ which we assign as the glass transition line. This is an arbitrary definition, but the behavior does not depend on the precise value chosen. This plot is for the high temperature situation ($T = T_f$). There are two well defined macrostates: a collapsed nonnative one and the native state. The line $Y = 0.1$ separates the two of them. Very importantly the

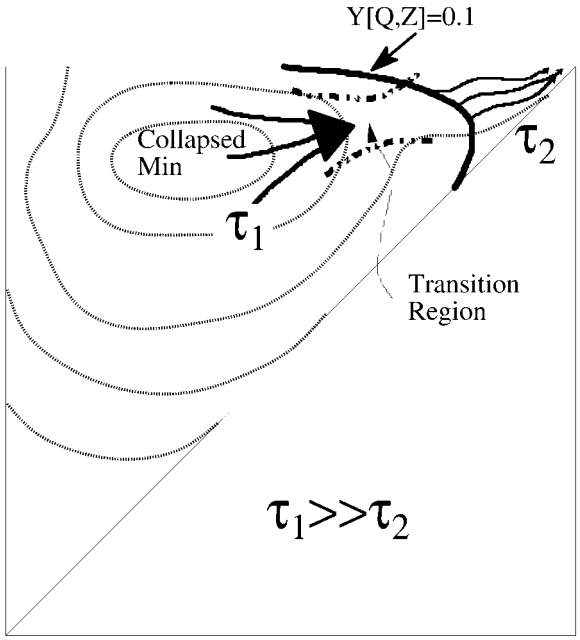

Fig. 12. Schematic drawing of free energy contours with the $Y = 0.1$ contour. The good (minimally frustrated) sequence at high hydrophobicity and high temperature ($T = T_f$). τ_1 is the time the chain spends diffusing about the molten globule minimum, searching for the transition region, while τ_2 is the time from the transition region to the final folded state (ignoring recrossings). For this sequence and these parameters $\tau_1 \gg \tau_2$. See Figure 14 for typical values.

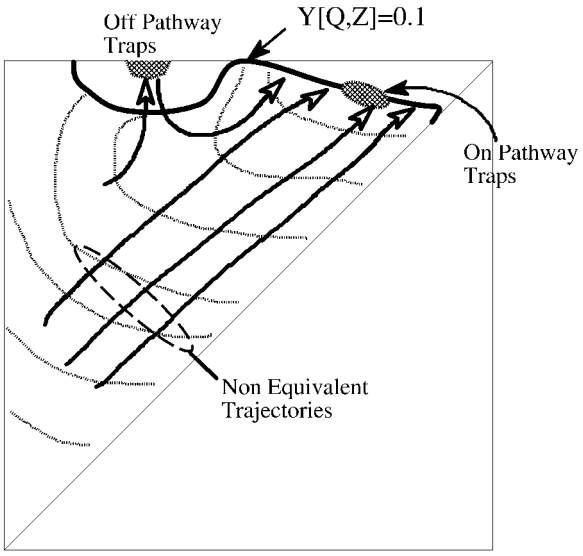

Fig. 13. Schematic drawing of free energy contours with the $Y = 0.1$ contour added. Good (minimally frustrated) sequence at high hydrophobicity at low temperature, below T_g. In this regime, folding is downhill. Different trajectories have different folding times, depending on the local roughness (diffusion coefficient) and whether traps are encountered or not. Both an off pathway and on pathway trap are shown.

transition region occurs before this line as folding proceeds, that is, on the side of the collapsed nonnative states. This picture is analogous to the one in the one dimensional funnel study (see figure 2 in ref. 10). The kinetics in this case is dominated by diffusive search for the transition region. Once the barrier has been crossed the folding is very rapid. So, even though there is a glass transition to a few states, those states are sufficiently connected to each other and to the folded state. The folding time is determined by the diffusion time to cross the barrier (which was studied in detail in ref. 13). This is an exponential process and the overall kinetics in this case is a simple exponential as we have seen in Figure 5.

Figure 13 shows the case at a low temperature (below T_g). Here the free energy profile is downhill, and there is no longer a well defined collapsed minimum or well defined transition region. The kinetics in this case is governed by diffusion down the free energy hill and can be fast or slow depending on the diffusion coefficient. At this temperature the diffusion coefficient is small leading to slow folding. More importantly the different routes down the hill may not be equivalent, that is, the time it takes to go down a specific path varies. Which route is taken depends on the random initial conditions and this gives rise to nonexponential kinetics as previously shown. In addition traps appear which may be either "on pathway" or "off pathway."

The glass transition line signals where the older notion of a "pathway" becomes appropriate. If it occurs late in the folding event, the slow dynamics of search may still be manageable. It is important to note T_g is actually a strong function of **Z** and **Q**. Thus there is no contradiction with the experimental observation that T_g of a folded proteins, that is, $\mathbf{Q} \approx 1$ (as studied by Frauenfelder[44–46]), is less than physiological temperatures. The transition that is important here is for **Z** comparable to the native but **Q** rather smaller. Also we must bear in mind that the Frauenfelder experiments reveal multiple tiers of substates, and that in the lower tiers certainly side chains are important in determining the glass transition. Thus, different tiers have different glass transition temperatures.

We now examine the combined free energy plus $Y[\mathbf{Q},\mathbf{Z}] = 0.1$ plots for our lattice model simulations. In addition to the free energy contours and the $Y = 0.1$ line, we also superimpose a given folding trajectory on each plot. Note, on each plot, we put just *one* folding trajectory not an ensemble average of trajectories. The trajectories were picked at random to illustrate various kinetic features. The trajectories should be represented as a connected line; however, that would quickly make the plot unreadable. Instead we place a dot for each time step at the appropriate **Q** and **Z** values. One last point; since

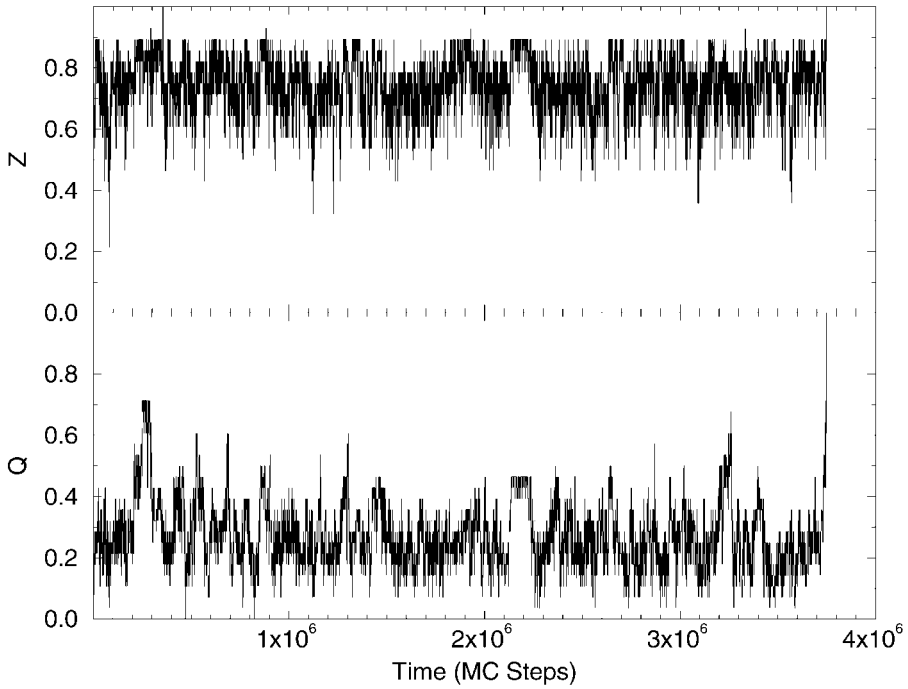

Fig. 14. Folding trajectory plotted as a function of time. These same data are plotted parametrically in Figure 15. This is the high hydrophobicity minimally frustrated (good) sequence at the folding temperature. The folding time for this run was approximately 3.7×10^6 MC steps. On the folding timescale, collapse to compact states is almost instantaneous. The chain then diffuses about the minimum of compact states ($\mathbf{Q} \approx 0.3$, $\mathbf{Z} \approx 0.8$) for most of the time, followed by a very short crossing of the transition region, and then the final folding.

each order parameter only has discrete values (separated by increments of 1/28) if we just plot the points they would all lie on a regular lattice and we could not tell in which region the trajectory is spending most time. So we add a random value to each dot to smear out the trajectory. This gives a visual representation in which the darkness of a region indicates how much time the trajectory spent there. Since the dots are unconnected the actual time course of the trajectory is not completely clear. However we can determine the exact time course by examining plots of **Q** and **Z** as a function of time (Fig. 14).

Figure 15 shows the folding trajectories for the good sequence at high hydrophobicity. The figure on the left illustrates trajectories at the folding temperature. The length of the trajectory shown is roughly equal to the mean folding time and is fairly typical of the trajectories under these conditions. The bulk of the time is spent diffusing about the collapsed nonnative minimum. The points have a roughly gaussian profile as would be expected. Once the transition region is crossed the trajectory rapidly moves to the folded state. Other trajectories (not shown here) show either multiple recrossing of the transition region or some transient trapping on the folding side of the barrier. But these trap states are extremely short lived and are a tiny fraction of the folding time. This figure shows a fairly typical two-state kinetic process. The figure on the right is for the same sequence and for the same value of κ but now at a very low temperature, below the glass transition temperature. Here the free-energy surface is downhill and there is now a free-energy minimum of collapsed nonnative states. The system still collapses rapidly but now, instead of diffusion about a free-energy minimum, it hops between discrete intermediates. This is because we are near the glass transition and the pathway language becomes appropriate. Note the trajectory appears to have several very dark regions and does not look at all gaussian. We do not expect a simple well-defined diffusion constant in this region. The trajectory could be described as a Lévy flight.† We also see two well defined traps: an off pathway one at $\mathbf{Q} \approx 0.2$ and an on pathway one at $\mathbf{Q} \approx 0.7$.

The next pair of plots (Fig. 16) is for the good sequence at low hydrophobicity. The temperature in the two cases is the same: an intermediate value below T_f but above T_g. The system is already multiexponential (see Fig. 6) at this temperature. The two

†A Lévy flight is a scale invariant random walk with no characteristic length for the jumps.[74–76] For these types of trajectories it is possible for the mean square displacement to increase superlinearly, that is, faster then normal (brownian) diffusion.

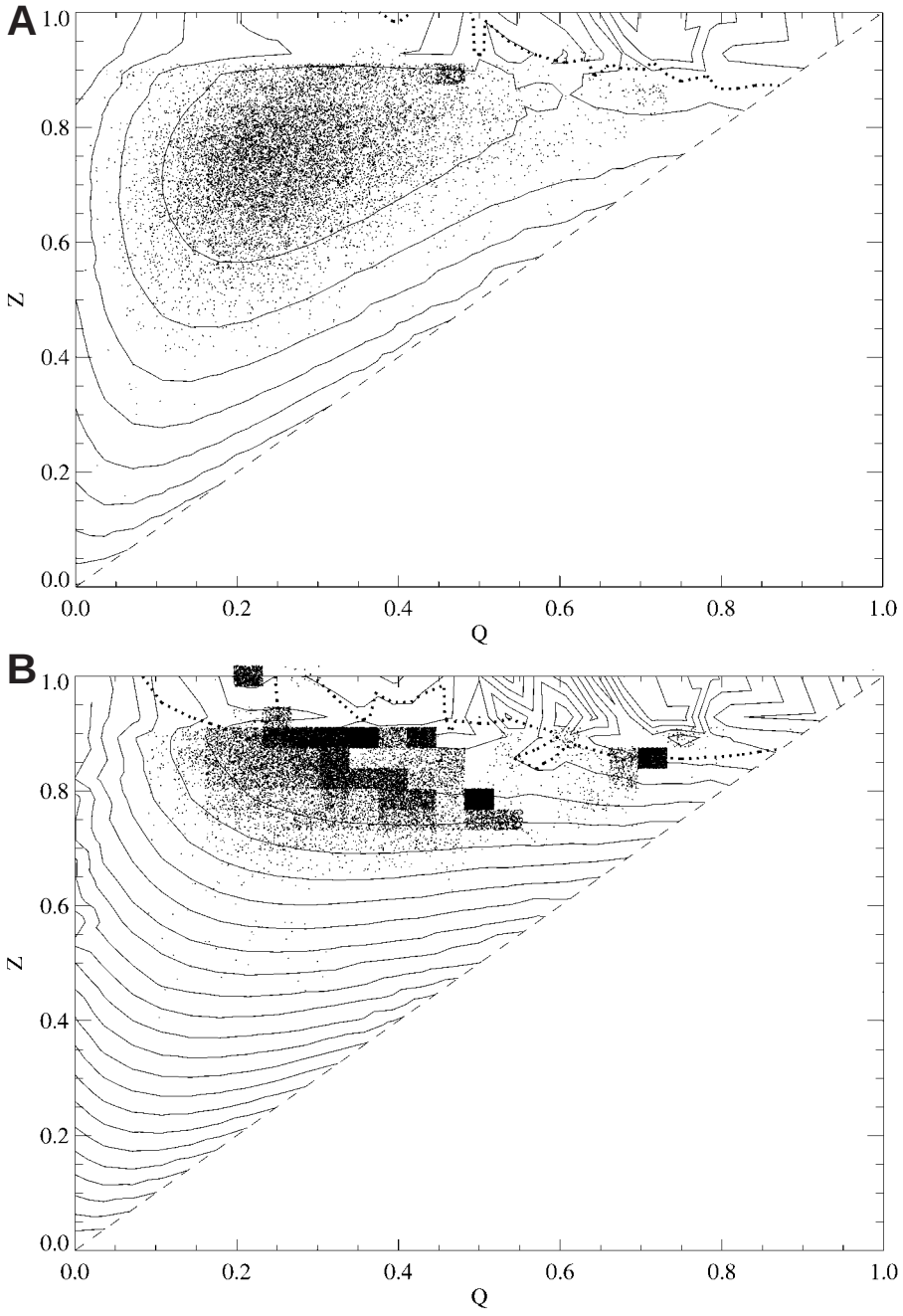

Fig. 15. Folding trajectory superimposed on free energy and Y[**Q**,**Z**] plots. The solid lines are the free energy contours spaced $2k_BT$ apart. The thick dotted line is Y[**Q**,**Z**] = 10^{-1}. The runs are for the good (minimally frustrated) sequence at high hydrophobicity. **A:** Folding temperature (1.509). **B:** $T = 0.89$ ($P_{nat} \approx 0.99$).

trajectories shown differ in elapsed time by two orders of magnitude (0.3×10^6 versus 30×10^6). The free energy surface is again downhill. The two trajectories take very different paths, each intersecting with the $Y = 0.1$ line at different points. The first hits the glass transition line at $\mathbf{Q} \approx 0.4$. It then uncompacts slightly and recrosses at a higher **Q** value. The second trajectory moves straight through at $\mathbf{Q} \approx 0.55$. The two pathways are nonequivalent with vastly different folding times giving rise to the nonexponential behavior in the folding.

The final trajectory plot (Fig. 17) is for the random (frustrated) sequence at high hydrophobicity and low temperature. Even at a very low temperature (at T_g) the system is still exponential. Looking at Figure 17 we see the free energy still has two well-defined

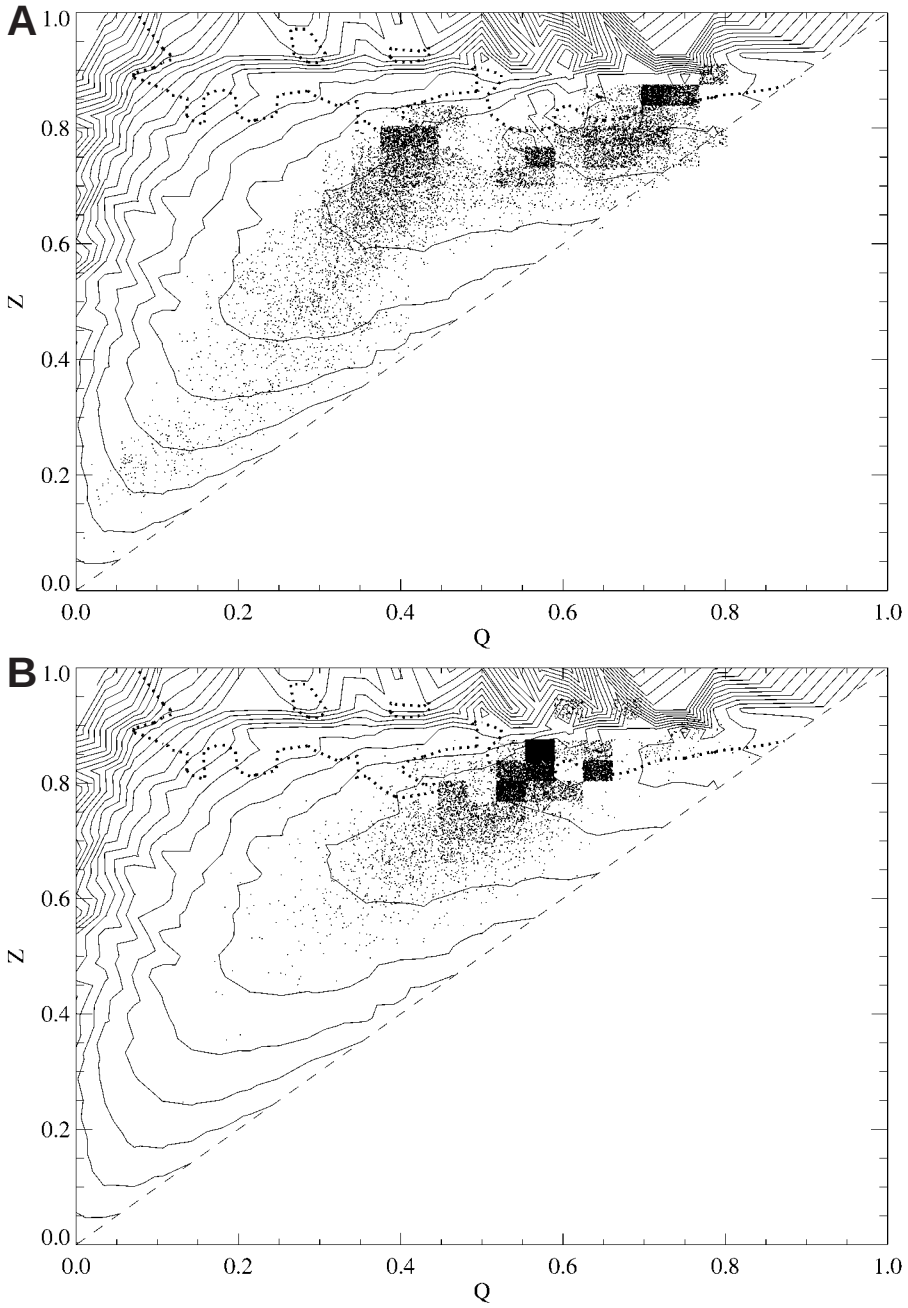

Fig. 16. Folding trajectory superimposed on free energy and Y[**Q**,**Z**] plots. Minimally frustrated (good) sequence at low hydrophobicity. Both plots are at the same temperature ($T = 1.41$, $P_{nat} = 0.89$). **A:** A short trajectory of 0.31×10^6 MC steps. **B:** 100 times longer (31×10^6 MC steps) than Figure A.

minima: a collapsed nonnative one and the native one. The $Y = 0.1$ line crosses right at the transition region so the system is neither clearly a type **IIa** or **IIb** (in the language of ref. 9) mechanism. We see that the trajectory is still dominated by diffusion about the collapse minimum. There are some well-defined traps but the timescale of escape from these traps is small compared to the folding time so the system is still single exponential.

CONCLUSION

The analysis presented in this report shows which global aspects of the energy landscape influence the "mechanisms" that would be ascribed to protein folding. The multidimensional funnel viewpoint suggests that, depending on the landscape characteristics, folding of small proteins can be a direct activated two step process or have "intermediates." The

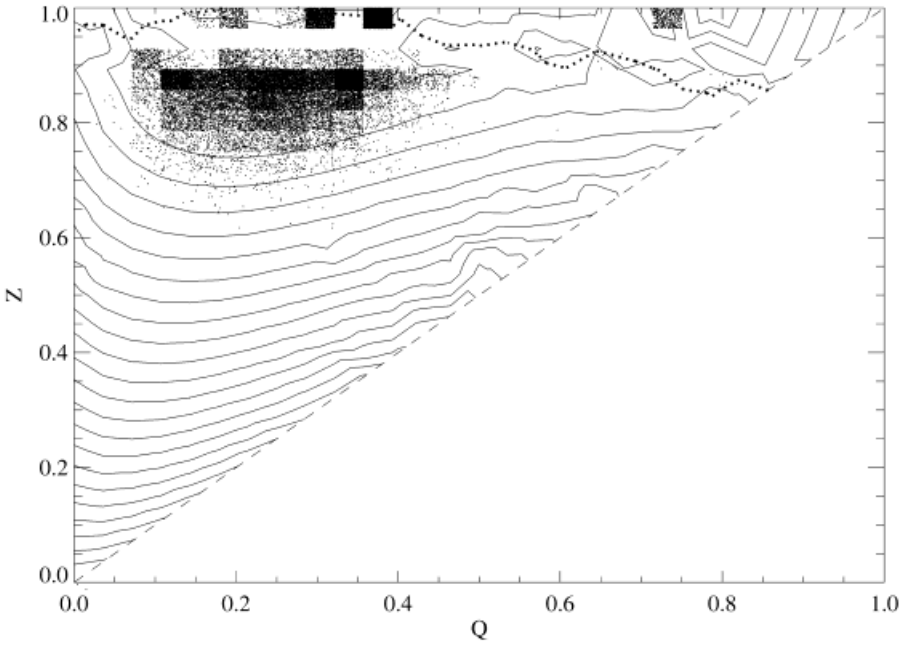

Fig. 17. Folding trajectory superimposed on free energy and Y[**Q**,**Z**] plots. Random (frustrated) sequence, high hydrophobicity, and low temperature ($T = 1$, $P_{nat} = 0.16$).

nature of the intermediates depends on the interplay between the ruggedness of the landscape reflected in the location of the glass transition and the overall funnel shape determining thermodynamic barriers. Because the landscape ruggedness and funnel geometry depend in a complex way on the order parameters, tuning a few parameters can utterly change the apparent mechanism while the basic structure of the landscape is only subtly varied. Thus quantifying or probing landscape characteristics, like the diffusion coefficient (which is related to the roughness), the drive towards non-specific collapse, or chain flexibility, provides an alternative goal to continued cataloging of intermediates for experimental studies. Recent work has already begun to examine some of these issues.[69,70,77,78] The different mechanisms we discussed used a model having two-order parameters but more have been invoked and are of interest. Most significantly the degree of local secondary structure formation can play a nearly independent role from **Q** and **Z** in characterizing the landscape properties. Real proteins may require explicit consideration of this dimension, when it is not slaved to collapse and native tertiary structure formation. Additionally, side-chain ordering may play a significant and possibly separable role from the collapse and topological ordering illustrated here. What is important, however, is that nothing new in terms of the qualitative type of mechanisms can arise from these effects. Instead the same kinetic scenarios for folding emerges. Nevertheless, it is important to understand these kinds of order if a truly microscopic interpretation of the landscape is to be achieved.

ACKNOWLEDGMENTS

This report was completed while P.G.W. was a Scholar in Residence at the Fogarty International Center for Advanced Study in the Health Sciences of the National Institute of Health (Bethesda, MD).

REFERENCES

1. Bryngelson, J.D., Wolynes, P.G. Spin glasses and the statistical mechanics of protein folding. Proc. Natl. Acad. Sci. U.S.A. 84:7524–7528, 1987.
2. Bryngelson, J.D., Wolynes, P.G. Intermediates and barrier crossing in a random energy model (with applications to protein folding). J. Phys. Chem. 93:6902–6915, 1989.
3. Bryngelson, J.D., Wolynes, P.G. A simple statistical field theory of heteropolymer collapse with application to protein folding. Biopolymers 30:177–188, 1990.
4. Leopold, P.E., Montal, M., Onuchic, J.N. Protein folding funnels: Kinetic pathways through compact conformational space. Proc. Natl. Acad. Sci. U.S.A. 89:8721–8725, 1992.
5. Chan, H.S., Dill, K.A. The protein folding problem. Phys. Today 46:24–32, 1993.
6. Chan, H.S., Dill, K.A. Energy landscapes and the collapse dynamics of homopolymers. J. Chem. Phys. 99:2116–2127, 1993.
7. Camacho, C.J., Thirumalai, D. Kinetics and thermodynamics of folding in model proteins. Proc. Natl. Acad. Sci. U.S.A. 90:6369–6372, 1993.
8. Abkevich, V.I., Gutin, A.M., Shakhnovich, E.I. Free energy landscape for protein folding kinetics: Intermediates, traps and multiple pathways in theory and lattice model simulations. J. Chem. Phys. 101:6052–6062, 1994.

9. Bryngelson, J.D., Onuchic, J.N., Socci, N.D., Wolynes, P.G. Funnels, pathways and the energy landscape of protein folding: A synthesis. Proteins 21:167–195, 1995.
10. Onuchic, J.N., Wolynes, P.G., Luthey-Schulten, Z., Socci, N.D. Towards an outline of the topography of a realistic protein folding funnel. Proc. Natl. Acad. Sci. U.S.A. 92:3626–3630, 1995.
11. Dill, K.A., Bromberg, S., Yue, K., Fiebig, K.M., Yee, D.P., Thomas, P.D., Chan, H.S. Principles of protein folding: A perspective from simple exact models. Protein Sci 4:561–602, 1995.
12. Boczko, E.M., Brooks, C.L. First-principles calculation of the folding free energy of a three-helix bundle protein. Science 269:393–396, 1995.
13. Socci, N.D., Onuchic, J.N., Wolynes, P.G. Diffusive dynamics of the reaction coordinate for protein folding funnels. J. Chem. Phys. 104:5860–5868, 1996.
14. Onuchic, J.N., Socci, N.D., Luthey-Schulten, Z., Wolynes, P.G. Protein folding funnels: The nature of the transition state ensemble. Folding Design 1:441–450, 1996.
15. Wolynes, P.G., Schulten, Z.L., Onuchic, J. Fast-folding experiments and the topography of protein folding energy landscapes. Chem Biol 3:425–432, 1996.
16. Mirny, L.A., Abkevich, V., Shakhnovich, E.I. Universality and diversity of the protein folding scenarios: A comprehensive analysis with the aid of a lattice model. Folding Design 1:103–116, 1996.
17. Dill, K.A., Chan, H.S. From Levinthal to pathways to funnels. Nature Struct. Biol. 4:10–19, 1997.
18. Daggett, V., Levitt, M. Protein unfolding pathways explored through molecular dynamics simulations. J. Mol. Biol. 232:600–619, 1993.
19. Kolinski, A., Skolnick, J. Monte Carlo simulations of protein folding. I. Lattice. model and interaction scheme. Proteins 18:338–352, 1994.
20. Kolinski, A., Skolnick, J. Monte Carlo simulations of protein folding. II. Application to protein A, ROP, and crambin. Proteins 18:353–366, 1994.
21. Šali, 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. 235:1614–1636, 1994.
22. Hao, M.-H., Scheraga, H.A. Monte Carlo simulation of a first-order transition for protein folding. J. Phys. Chem. 98:4940–4948, 1994.
23. Chan, H.S., Dill, K.A. Protein folding in the landscape perspective. Proteins 30:2–33, 1998.
24. Govindarajan, S., Goldstein, R.A. The foldability landscape of model proteins. Biopolymers 42:427–438, 1997.
25. Khorasanizadeh, S., Peters, I.D., Roder, H. Evidence for a three-state model of protein folding from kinetic analysis of ubiquitin variants with altered core residues. Nature Struct. Biol. 2:193–205, 1995.
26. Guo, Z.Y., Thirumalai, D. Kinetics of protein folding: Nucleation mechanism, time scales, and pathways. Biopolymers 36:83–102, 1995.
27. Thirumalai, D., Guo, Z.Y. Nucleation mechanism for protein folding and theoretical predictions for hydrogen-exchange labeling experiments. Biopolymers 35:137–140, 1995.
28. Plotkin, S.S., Wang, J., Wolynes, P.G. Statistical mechanics of a correlated energy landscape model for protein folding funnels. J. Chem. Phys. 106:2932, 1997.
29. Socci, N.D., Onuchic, J.N. Kinetic and thermodynamic analysis of proteinlike heteropolymers: Monte Carlo histogram technique. J. Chem. Phys. 103:4732–4744, 1995.
30. Socci, N.D., Onuchic, J.N. Folding kinetics of proteinlike heteropolymers. J. Chem. Phys. 101:1519–1528, 1994.
31. Kauzmann, W. Adv. Protein Chem. 14:1–64, 1959.
32. Dill, K.A. Dominant forces in protein folding. Biochemistry 29:7133–7155, 1990.
33. Chan, C.-K., Hu, Y., Takahashi, S., Rousseau, D.L., Eaton, W.A., Hofricher, J. Submillisecond protein folding kinetics studied by ultrarapid mixing. Proc. Natl. Acad. Sci. U.S.A. 94:1779–1784, 1997.
34. Ballew, R.M., Sabelko, J., Gruebele, M. Observation of distinct nanosecond and microsecond protein folding events. Nature Struct. Biol. 3:923–926, 1996.
35. Sosnick, T.R., Mayne, L., Hiller, R., Englander, S.W. The barriers in protein folding. Nature Struct. Biol. 1:149–156, 1994.
36. Lattman, E.E., Rose, G.D. Protein folding-what's the question. Proc. Natl. Acad. Sci. U.S.A. 90:439–441, 1993.
37. Sheinerman, F.B., Brooks, C.L. Molecular picture of folding of a small α/β protein. Proc. Natl. Acad. Sci. U.S.A., submitted, 1997.
38. Guo, Z., Brooks, C.L., Boczko, E.M. Exploring the folding free energy surface of a three helix bundle protein. Proc. Natl. Acad. Sci. U.S.A. 94:10161–10166, 1997.
39. Guo, Z., Brooks, C.L. Thermodynamics of protein folding: A statistical mechanical study of a small all β protein. Biopolymers, submitted, 1997.
40. Luthey-Schulten, Z., Ramirez, B.E., Wolynes, P.G. Helix-coil, liquid crystal and spin glass transitions of a collapsed heteropolymer. J. Phys. Chem. 99:2177–2185, 1995.
41. Uversky, V.N., Ptitsyn, O.B. All-or-none solvent-induced transitions between native, molten globule and unfolded states in globular proteins. Folding Design 1:117–122, 1996.
42. Wang, J., Onuchic, J., Wolynes, P. Statistics of kinetic pathways on biased rough energy landscapes with applications to protein folding. Phys. Rev. Lett. 76:4861–4864, 1996.
43. Leite, V.B.P., Onuchic, J.N. Structure and dynamics of solvent landscapes in charge-transfer reactions. J. Phys. Chem. 100:7680–7690, 1996.
44. Iben, I., Braunstein, D., Doster, W., et al. Glass behavior of a protein. Phys. Rev. Lett. 62:1916–1919, 1989.
45. Tilton, R., Dewan, J., Petsko, G. Effects of temperature on protein structure and dynamics: X-ray crystallographic studies of the protein ribonuclease-A at nine different temperatures from 98 to 320 K. Biochemistry 31:2469–2481, 1992.
46. Ansari, A., Jones, C.M., Henry, E.R., Hofrichter, J., Eaton, W.A. The role of solvent viscosity in the dynamics of protein conformational changes. Science 256:1796–1798, 1992.
47. Angell, C.A. Formation of glasses from liquids and biopolymers. Science 267:1924–1935, 1995.
48. Sochava, I.V., Smirnova, O.I. Heat capacity of hydrated and dehydrated globular proteins. Food Hydrocolloids 6:513–524, 1993.
49. Miller, R., Danko, C.A., Fasolka, M.J., Balazs, A.C., Chan, H.S., Dill, K.A. Folding kinetics of proteins and copolymers. J. Chem. Phys. 96:768–780, 1992.
50. Chan, H.S., Dill, K.A. Transition states and folding dynamics of proteins and heteropolymers. J. Chem. Phys. 100: 9238–9257, 1994.
51. Plotkin, S.S., Wang, J., Wolynes, P.G. Correlated energy landscape model for finite, random heteropolymers. Phys. Rev. E 53:6271–6296, 1996.
52. Takada, S., Wolynes, P.G. Statics, metastable states and barriers in protein folding: A replica variational approach. Phys. Rev. E 55:4562, 1997.
53. Plotkin, S.S., Wang, J., Wolynes, P.G. Configurational diffusion on a locally connected correlated energy landscape: Application to finite random heteropolymers. J. Phys. I (France) 7:395, 1997.
54. Roan, J., Shakhnovich, E. Dynamics of heteropolymers in dilute solution: Effective equation of motion and relaxation spectrum. Phys. Rev. E 5:5340–5357, 1996.
55. Thirumalai, D., Ashwin, V., Bhattacharjee, J.K. Dynamics of random hydrophobic-hydrophilic copolymers. Phys. Rev. Lett. 77:5385–5388, 1996.
56. Timoshenko, E.G., Kuznetsov, Y.A., Dawson, K.A. Kinetics of a gaussian random copolymer as a prototype for protein folding. Phys. Rev. E 54:4071–4086, 1996.
57. Takada, S., Portman, J.J., Wolynes, P.G. An elementary mode coupling theory of random heteropolymer dynamics. Proc. Natl. Acad. Sci. U.S.A. 94:2318–2321, 1997.
58. Austin, R.H., Beeson, K.W., Eisenstein, L., Frauenfelder,

H., Gunsalus, I.C. Dynamics of ligand binding to myoglobin. Biochemistry 14:5355–5373, 1975.
59. Frauenfelder, H., Parak, F., Young, R.D. Conformational substates in proteins. Annu. Rev. Biophys. Biophys. Chem. 17:451–479, 1988.
60. Steinbach, P.J., Chu, K., Frauenfelder, H., Johnson, J.B., Lamb, D.C., Nienhaus, G.U., Young, R.D. Determination of rate distributions from kinetic experiments. Biophys. J. 61:235–245, 1992.
61. Levinthal, C. Are there pathways for protein folding? J. Chim. Phys. 65:44–45, 1968.
62. Levinthal, C. How to fold graciously. In: "Mossbauer Spectroscopy in Biological Systems." DeBrunner, P., Tsibris, J., Munck, E. (eds.). Urbana, IL: University of Illinois Press, 1969:22–24.
63. Betancourt, M.R., Onuchic, J.N. Kinetics of proteinlike models: The energy landscape factors that determine folding. J. Chem. Phys. 103:773–787, 1995.
64. Doi, M., Edwards, S.F. "The Theory of Polymer Dynamics." Oxford, U.K.: Oxford University Press, 1986.
65. Jackson, S.E., Fersht, A.R. Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. Biochemistry 30:10428–10435, 1991.
66. Huang, G.S., Oas, T.G. Submillisecond folding of monomeric λ repressor, Proc. Natl. Acad. Sci. U.S.A. 92:6878–6882, 1995.
67. Pascher, T., Chesick, J.P., Winkler, J.R., Gray, H.B. Protein folding triggered by electron transfer. Science 271:1558–1560, 1996.
68. Robinson, C.R., Sauer, R.T. Equilibrium stability and submillisecond refolding of a designed single-chain arc repressor. Biochemistry 35:13878–13884, 1996.
69. Mines, G.A., Pascher, T., Lee, S.C., Winkler, J.R., Gray, H.B. Cytochrome c folding triggered by electron transfer. Chem Biol 3:491–497, 1996.
70. Scalley, M.L., Baker, D. Protein folding kinetics exhibit an Arrhenius temperature dependence when corrected for the temperature dependence of protein stability. Proc. Natl. Acad. Sci. U.S.A. 94:10636–10640, 1997.
71. Jennings, P.A., Finn, B.E., Jones, B.E., Matthews, C.R. A reexamination of the folding mechanism of dihydrofolate reductase from *Escherichia coli:* Verification and refinement of a four-channel model. Biochemistry 32:3783–3789, 1995.
72. Iwakura, M., Jones, B.E., Falzone, C.J., Matthews, C.R. Collapse of parallel folding channels in dihydrofolate reductase from *Escherichia coli* by site-directed mutagenesis. Biochemistry 32:13566–13574, 1993.
73. Shakhnovich, E.I., Gutin, A.M. Implications of thermodynamics of protein folding for evolution of primary sequences. Nature 346:773–445, 1990.
74. Mandelbrot, B. "The Fractal Geometry of Nature." New York: Freeman, 1982.
75. Shlesinger, M., Zaslavsky, G., Frisch, U. "Lévy Flights and Related Topics in Physics." New York: Springer-Verlag, 1995.
76. Klafter, J., Shlesinger, M.F., Zumofen, G. Beyond brownian motion. Phys. Today 49:33–39, 1996.
77. Burton, R.E., Huang, G.S., Daugherty, M.A., Calderone, T.L., Oas, T.G. The energy landscape of a fast-folding protein mapped by Ala → Gly substitutions. Nature Struct. Biol. 4:305–310, 1997.
78. Beeser, S.A., Goldenberg, D.P., Oas, T.G. Enhanced protein flexibility caused by a destabilizing amino acid replacement in BPTI. J. Mol. Biol. 269:154–164, 1997.