PROTEINS: Structure, Function, and Genetics 30:2–33 (1998)

Protein Folding in the Landscape Perspective: Chevron Plots and Non-Arrhenius Kinetics

Hue Sun Chan and Ken A. Dill*
Department of Pharmaceutical Chemistry, University of California, San Francisco, California

ABSTRACT **We use two simple models and the energy landscape perspective to study protein folding kinetics. A major challenge has been to use the landscape perspective to interpret experimental data, which requires ensemble averaging over the microscopic trajectories usually observed in such models. Here, because of the simplicity of the model, this can be achieved. The kinetics of protein folding falls into two classes: multiple-exponential and two-state (single-exponential) kinetics. Experiments show that two-state relaxation times have "chevron plot" dependences on denaturant and non-Arrhenius dependences on temperature. We find that HP and HP+ models can account for these behaviors. The HP model often gives bumpy landscapes with many kinetic traps and multiple-exponental behavior, whereas the HP+ model gives more smooth funnels and two-state behavior. Multiple-exponential kinetics often involves fast collapse into kinetic traps and slower barrier climbing out of the traps. Two-state kinetics often involves entropic barriers where conformational searching limits the folding speed. Transition states and activation barriers need not define a single conformation; they can involve a broad ensemble of the conformations searched on the way to the native state. We find that unfolding is not always a direct reversal of the folding process. Proteins 30:2–33, 1998.**

Key words: chevron plot; energy landscape; folding funnel; kinetic trap; lattice models; non-Arrhenius behavior

TWO PERSPECTIVES: ENERGY LANDSCAPES VS. FOLDING PATHWAYS

Our aim here is to show that the energy landscape perspective of protein folding, and some simple models on which it is based, can explain some of the main classes of experimental folding data—single- and multiple-exponential relaxations, chevron plots, and non-Arrhenius kinetics. Recently, a so-called "New View"[1,2] of protein folding kinetics has emerged. The New View replaces the concept of "folding pathway" with "energy landscapes."[3–8] Anfinsen[9,10] and others showed that proteins can fold reversibly to stable states, which was taken as evidence that folding is "path independent." In contrast, Levinthal[11,12] argued that proteins would fold too slowly by undirected random searching of conformations. So he concluded that protein folding must follow a "pathway," which is "a well-defined sequence of events which follow one another so as to carry the protein from the unfolded random coil to a uniquely folded metastable state."[11] These two lines of argument have led to the "Levinthal paradox"[13,14]: How could folding be pathway dependent and pathway independent at the same time?

But according to the energy landscape perspective, there is no paradox because the new view recognizes that "folding pathways" are not the correct solution to the kinetic problem Levinthal posed.[8] The landscape perspective readily explains the process of reaching a global minimum in free energy (satisfying Anfinsen's experiments) and doing so quickly (satisfying Levinthal's concerns) by multiple folding routes on funnel-like energy landscapes[15,16] (reviewed in Refs. 3–8). Instead of viewing folding as a process in which all chains perform essentially the same sequence of events to reach the native state, the new view envisions folding as representing the ensemble average of a process that is microscopically more heterogeneous. In this process, each individual protein molecule may follow its own trajectory, but just like skiers down a mountain, they all may eventually reach the same point at the bottom, the native state.

How might experiments distinguish pathways from more complex landscapes? Experiments in the 1970s began to explore folding pathways.[17–24] But the pathways that are found by experimentalists are very different from the microscopic pathways defined by Levinthal. The main observables in experiments have been exponential decays of optical signals. Mass-action models are used to fit the time constants

Contract grant sponsor: NIH.
Parts of this paper were given as the First Anfinsen Memorial Lecture by K.A.D. at the Johns Hopkins Protein Folding Meeting, Coolfont, West Virginia, March 1996. Electronic version of the energy landscape drawings in this paper are available at the website http://laplace.ucsf.edu. H.S.C.'s E-mail: chan@maxwell.ucsf.edu
*Correspondence to: Ken A. Dill, Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94143-1204.
E-mail: dill@maxwell.ucsf.edu
Received 11 March 1997; Accepted 30 June 1997

and amplitudes by postulating macroscopic states, such as unfolded, native, and various "intermediate" states. The "pathways" defined by the arrows that connect such states in mass-action kinetics laws describe ensemble averages that do not bear on whether individual chains perform the same sequence of events.

Statistical mechanical models of protein folding[13,25–29] differ from the mass-action kinetics models in two main respects. First, the statistical mechanical models attempt to represent—albeit at low resolution and based on several approximations—the chain and its microscopic interactions. Second, the statistical mechanical models attempt to capture the ensemble nature of the chain conformations. Hence, they can help to relate the microscopic chain dynamics to the macroscopic experimental observables.

What Are the Axes and Coordinates on Energy Landscapes?

The energy landscape figures shown in this paper have two lateral coordinates, say θ_1 and θ_2, and one vertical coordinate, which is the "internal free energy" $E(\theta_1, \theta_2)$ of a chain conformation. The lateral coordinates represent degrees of freedom of the chain, such as the torsion angles of the backbone and side chains. For real and model proteins, there are, in fact, tens to thousands of lateral coordinates, θ_1, θ_2, θ_3, . . ., θ_m, where $m \gg 2$, but limitations of the printed page make it impossible to show the energy as a function of more than two. Hence, only three-dimensional figures of energy landscapes are shown here as artist's renditions. The real energy landscape is a high-dimensional surface defined by the function $E(\theta_1, \theta_2, \theta_3, \ldots, \theta_m)$.

What is the meaning of the internal free energy E? E represents the sum of all the intrachain enthalpies and solvent interactions (hydrogen bonding, torsion angle energies, monomer contacts, etc.) of a molecule in a given single conformation, defined by $(\theta_1, \theta_2, \theta_3, \ldots, \theta_m)$. Such a quantity is not just an energy or enthalpy because it contains solvation components, which can involve entropies due to water ordering. Hence E is a free energy. But E is not the macroscopic free energy that would be measured in a folding experiment, because E describes only a single-chain conformation, not the ensemble average over all chain conformations. Hence the term *internal* in describing this free energy, which means that it pertains only to a single chain conformation. The internal free energy contains no accounting of the chain entropy, i.e., of how many different chain conformations have a given value of E.

Why should we divide up the total free energy into an internal free energy and a chain entropy, instead of, say, a sum of all enthalpies and a sum of all entropies? The concept of the internal free energy has a long history that goes back to Flory-Huggins theory, developed in the early 1940s (see review in Ref. 30). In the polymer literature, the internal free energy is described by the so-called Flory χ parameter.[30] The virtue of the internal free energy is that it can be measured in independent experiments, such as in oil/water partitioning or vaporization experiments (Ref. 31 and references therein). In contrast, the chain entropies are not directly accessible by experiments. The idea is to use measurable data on interactions among individual monomer units to predict polymer conformational changes. This is the basis for energy landscapes as internal free energies vs. conformational degrees of freedom.

An implication of the use of the internal free energy as the vertical coordinate is that temperature or denaturants can stretch or compress the landscape in the vertical direction. At high enough temperatures or denaturing concentrations of denaturants, there is little driving force to fold, which means the landscape is quite flat—the native structure is not much lower in internal free energy than other conformations. As temperature (or denaturant concentration) is lowered, the hydrophobic interaction strengthens, which is equivalent to "pulling down" on the native state to stretch the landscape vertically, and now the native state has a deeper well relative to other conformations. The slopes of landscapes play a role in folding kinetics; hence, conditions that are "more native" can also lead to faster folding.

The many lateral coordinates are the microscopic degrees of freedom. Each conformation is a point on the landscape. Lateral distances must correspond to some properly defined kinetic distances if the geometric shape of the energy landscape is to provide an intuitive depiction of the kinetics. This means that conformations that easily interconvert into one another should have similar lateral coordinates.

Microscopic models made some general predictions about folding kinetics. Such models indicate that kinetic traps slow down folding. Folding intermediates are transiently trapped conformations.[3,4,7,16,26,32–47] The temperature dependences of folding rates have been modeled on the basis of assuming temperature-independent intrachain interactions.[3,44,48] A specific chain structure or "nucleus" has been proposed as a model for the kinetic transition state.[42,43,49,50] Attempts have been made to generalize short-chain lattice results by scaling arguments.[6,51,52]

A shortcoming of the new models has been the shortage of predictions of experimental observables. This is in part because simulations often follow only a few folding trajectories, rather than the ensemble-averaged quantities observed in experiments. Here, we study ensemble averages. Current understanding of two-state, fast-folding kinetics relies heavily on experimental chevron plots of rate constants vs. denaturant,[53–62] their interpretation in terms of one-dimensional energy profiles and reaction diagrams and transition states,[58,61,63–75] and non-Arrhenius kinetics of rate constants vs. temperature.[54,58,61,65,70,73,76,77]

Here, we describe two simple statistical mechanical models and show how these experimental behaviors can be rationalized by the energy landscape perspective.

TWO MODELS OF PROTEIN FOLDING KINETICS

HP Model

We model proteins as two-dimensional (2D), self-avoiding walk chains of n monomers on the square lattice. One model we use is the HP model of two monomer types: H for hydrophobic and P for polar monomers. In the HP model, each lattice nearest-neighbor HH contact interaction has energy $\epsilon \leq 0$, whereas all other contact types (HP, PP) are neutral, i.e., have zero energy.[4,29,78–80] Native states are determined by exhaustive enumeration of the conformations. We consider here only sequences that have a single lowest energy conformation, based on the view that they are most representative of real proteins. Kinetics of the 2D HP model have been studied by using Monte Carlo simulations[33,35] and by analytical transition matrix methods.[37,38,81]

For most HP sequences, folding involves multiphasic kinetics, having at least two stages: a fast collapse to compact conformations and a slow barrier-climbing process to escape traps and reattempt to fold.[4,7,33,35,38] These kinetic features resemble those of other simple models.[16,32,34–36,39–47,82] The landscape perspective suggests that the catalytic action of chaperonins may involve binding to misfolded kinetic intermediates with exposed nonpolar surfaces,[83,84] pulling the misfolded proteins apart nonspecifically, and allowing them to try again to refold (Ref. 81 and references therein).

The virtues of the short-chain HP model are: 1) that it can provide a *complete* characterization of the model energy landscape based on simple rules of viable chain movements[37,38]; 2) it captures what we believe are the dominant interactions in folding, and 3) it gives physical insights that can be tested by experiments or by more sophisticated models. The general features of protein folding kinetics appear to be captured in 2D models.[33,35,37,38,81,85–88] Analytical approximations[37,38,81] provide a fast way to compute behavior over much longer times than is feasible in Monte Carlo simulations and to obtain approximate ensemble averages. We can compare kinetics of sequences whose folding times differ by many orders of magnitudes.

HP+ model

Most HP model sequences have rugged energy landscapes with many kinetic traps. But some real proteins appear to have smoother landscapes—meaning that they have few traps, resulting in fast folding and two-state relaxations.[53–59,61,62] To model those, we introduce a variant of the HP model we call the HP+ model. It is now well established that adding native propensities helps model proteins fold faster.[82,85–87,89–97] Adding repulsive and unfavorable interactions to model energy matrices also tends to decrease sequence degeneracy and increase structural encodability.[98,99]

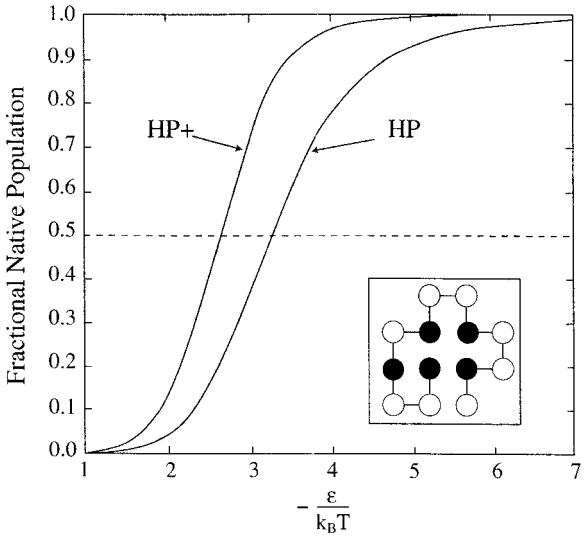

Fig. 1. Melting curves of a 2D lattice model protein (given in its unique native structure in the inset; H, filled circles; P, open circles) using the HP and HP+ energy models. Fractional native population vs. the "sticking strength" $\epsilon/(k_B T)$. On the basis of data given in Figure 2, native populations are obtained by Boltzmann averaging.[38,157]

For an HP sequence with a unique native structure,[38,79] the HP+ energy is defined by $\epsilon < 0$ for each native HH contact, $\epsilon = 0$ for each native HP or PP contact, and $-\epsilon$ for all nonnative contacts ($-\epsilon > 0$). It follows that the HP and its corresponding HP+ sequences have the same unique native structure. The only difference is the unfavorable energies for nonnative contacts in the HP+ energy function. This extra interaction in the HP+ model is intended to capture—in a very simple way, and without additional parameters—more energetic specificity than is in the HP model.

The HP+ model is similar to the "strong specificity limit" of the model of Taketomi et al.,[85] sometimes called the "Gō model,"[85–87,89] except that nonnative interactions that are neutral in the Gō model are now unfavorable in the HP+ model. The favorable and unfavorable interactions are equal in magnitude in the HP+ model. Our aim here is to explore principles, not to model real amino acid interactions accurately. Real proteins have interactions more specific than that modeled by the HP model; the HP+ model captures them in a crude way. Our goal here is simply to use the comparison between the HP and HP+ sequences to bracket behaviors intermediate between the two extremes of nonspecific and specific interactions. Models with high specificity can also be obtained by using multiple-letter codes and sequence design (see, for example, Refs. 94, 99–102).

Figure 1 shows the model sequence used to illustrate most of the results below. It is typical of others. The sequence is shown in its unique native structure for either the HP or HP+ energy function. Figure 1 shows the "melting curves," which characterize the thermodynamics of the models. The melting curves are computed by exact enumeration of the partition functions. The chain has a greater native stability at any HH sticking strength $\epsilon/(k_B T) < 0$—and a sharper transition—with the HP+ energy function than with the HP energy function. (Here k_B is Boltzmann's constant and T is absolute temperature.) The equilibrium transition midpoints (50% native) of the HP and HP+ chains are at $\epsilon = -3.25 k_B T$ and $-2.61 k_B T$, respectively. The corresponding interaction strengths necessary for reaching 98% equilibrium native population are $\epsilon = -6.16 k_B T$ and $-4.17 k_B T$.

Figure 2 shows the energy ladders of the chain using the two different energy functions. At each energy level, the total number of conformations is given by the numbers in the ovals and the squares. Kinetic properties are calculated by using move set MS2 defined elsewhere.[37,38,81] Figure 2 shows that the two main differences in the energy landscapes are that there are more hills on the HP+ landscape (positive energy levels) and that there are fewer kinetic traps on the HP+ landscape (conformation counts shown in the squares). The HP+ chain is more stable than the same sequence in the HP energy function because many nonnative conformations in HP+ are more unstable, due to the replacement of some neutral or favorable interactions in the HP model with unfavorable ones in the HP+ model. Many of the compact nonnative conformations are strongly disfavored by the HP+ potential. The HP+ chain has no conformations with $E = -4$ (the native energy is $E = -5$), so it has a larger energy gap[38,100–103] than the HP chain.

Because the HP+ chain has fewer trapped conformations than the HP chain (24 vs. 88), the HP+ chain folds faster. The arrows in Figure 2 show that many conformations that are kinetic traps when the chain is subject to the HP potential become throughway (nontrapped) conformations[38] in the HP+ potential. Gutin et al.[94] too have found that adding repulsions to a different model also speeds folding.

What are kinetic traps in these models? Traps are low-energy conformations (i.e., having many HH contacts) and therefore are usually quite compact.[37,38] But to escape traps by breaking wrong contacts and climbing over energy barriers, chains are often hindered by the severe steric constraints inside the compact chain. Hence, the repulsions inherent in the HP+ energy function can destabilize these trapped conformations. Thus, the stability and folding rate of a model protein can be increased not only by reducing the conflict among different interactions (i.e., minimizing frustration) in the native state[3,13,25,104] but also by increasing the frustration of the compact denatured states.

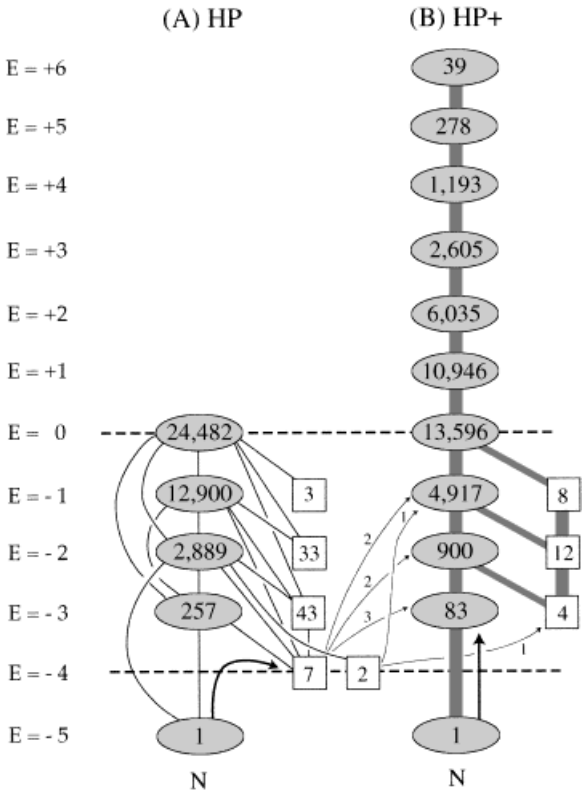

Fig. 2. Conformational flow diagrams for the sequence in Figure 1, in the HP and HP+ models: numbers of conformations at each energy level. The HP+ model has many high-energy conformations and few traps. The energy E of the conformational states is given in units of $|\epsilon|$. Conformations are also classified by d, which is the minimum number of uphill steps a conformation must take before it can reached the native state, hence $d > 0$ conformations are traps.[37,38] Lines show the kinetic accessibilities of the (E, d) states. In ovals: numbers of conformations along "throughway" paths (i.e., involving no uphill steps, $d = 0$). In squares: numbers of trapped conformations that must first go uphill before reaching the native state. First column of squares indicates at least one step uphill ($d = 1$); second column of squares indicates at least two steps uphill ($d = 2$). A line between two states of the HP sequence (**A**) indicates that at least a pair of conformations, one from each state, is kinetically adjacent. For the HP+ sequence in (**B**), adjacencies are too numerous to show in their entirety. A thick line in (B) represents kinetic adjacency between two states, whereas direct adjacencies between two states that are connected via other states are not shown. Arrows from the $E = -4$ kinetic traps of the HP sequence in the left diagram indicate the HP+ conformational states to which these conformations belong. Number of corresponding conformations are given along arrows. Arrows from the native (N) indicate the dominant unfolding paths when conditions are favorable for folding.

What Is the Reaction Coordinate? Kinetic vs. Thermodynamic Reaction Coordinates

To model the kinetic progress in folding, it is helpful to define a reaction coordinate. In this section, we describe the two different types of reaction coordinates, which we call thermodynamic and kinetic reaction coordinates. We prefer kinetic reaction coordinates. The differences are substantial and even affect the signs of some qualitative predictions.

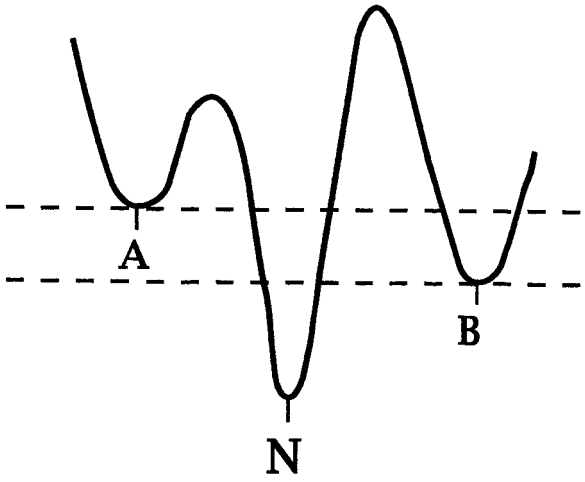

Fig. 3. Defining thermodynamic vs. kinetic reaction coordinates. Which state, A or B, is further "along the reaction coordinate" toward the native state N? State B is *energetically* closer to N (lower energy), but A is *kinetically* closer (smaller barrier to cross).

To define kinetic progress requires a definition for "closeness to the native structure." A thermodynamic reaction coordinate defines closeness to the native structure in terms of the *energy* of the conformation, whereas a kinetic reaction coordinate defines closeness to the native structure in terms of how quickly that conformation can transform to the native state. Perhaps the most common thermodynamic reaction coordinate used in model studies is Q, the fraction of native contacts that are present in a conformation.[6,40,101,105] Although thermodynamic reaction coordinates are adequate for describing kinetics on relatively smooth funnel-like landscapes, they are not well suited for kinetics on rugged landscapes that are dominated by kinetic traps.[106]

Figure 3 shows the problem. A thermodynamic reaction coordinate sees some deeply trapped conformations (B) as being "nearly native" (because B has low energy), even though such conformations must overcome high-energy barriers to reach the native state. But a kinetic progress coordinate should describe, at least at some rudimentary level, the fraction of *time* that has elapsed, or that remains, for the folding, rather than the fraction of *energy* that remains. By using a thermodynamic reaction coordinate, B in Figure 3 is closer to native N than A is. But by using a kinetic reaction coordinate, A is closer to N than B. For landscapes with kinetic traps, thermodynamic reaction coordinates do not characterize well the kinetics, because they completely neglect energy barriers.

To define a kinetic reaction coordinate in lattice models requires a "move set," which prescribes the number of moves that are required for one conformation to reach another.[37,38] Any conformation that can reach another by only a single move in the move set is called *adjacent*, or "directly connected" in a kinetic sense. In one time step of the simulation, a conformation can make transitions only to an adjacent conformation.

Here, we define folding progress by using a kinetic reaction coordinate, δ.[38] For a given move set (MS2 is used here, see Refs. 37 and 38 for definition), the distance δ from any conformation to the native state is defined as the minimum total number of steps necessary to reach the native structure along a path that is least energetically costly, i.e., a path that requires a minimum total energy in breaking favorable contacts or forming unfavorable contacts.[37,38] δ is similar to a parameter suggested by Kuntz et al.[107] for conformational comparison. By definition, a single kinetic move can only change δ by either 0 or ± 1. We use δ rather than Q or other thermodynamic order parameters for kinetic analyses because δ represents *kinetically achievable progress* toward the native state.[4,38] Appendix A discusses the relation between δ and Q.

Here is a practical distinction between thermodynamic and kinetic reaction coordinates. By using a typical thermodynamic reaction coordinate, as the internal free energy of a chain proceeds downhill through a series of conformations, folding is seen to always make *forward progress toward* the native state. In contrast, by using a kinetic reaction coordinate, folding of denatured chains is sometimes seen to be a process of *first moving away from the native state* as the chains fall into traps and then *slowly proceeding back toward the native state* as the chains exit the traps and become more native-like. The reason the chains start by moving away from the native state is because kinetically trapped chains must cross higher barriers than unfolded chains must cross[38] and are therefore kinetically more distant from the native state. In kinetic terms, falling into traps is not "progress" toward the native state; it is a hindrance to progress.

Figure 4 shows how progress on bumpy landscapes with many traps (HP) is first backward then forward, whereas progress on smooth funnels (HP+) is mostly forward. Figure 4 shows ensemble-averaged properties as a function of position along the kinetic reaction coordinate. When conditions are changed from denaturating to native, folding begins with many conformations (i.e., having a low value of $-1n\ \Omega$) that are open (small ρ) at $\delta \approx 4$–7, which is the average kinetic distance of denatured conformations from the native structure. The E curves show the average energetic drive. For the HP energy function (Fig. 4A), some of the model molecules will reach lower energies by moving left (i.e., folding quickly along energetically downhill "throughway paths"), whereas others will move right (become compact, high ρ, and low E) and become trapped. The trapped conformations then climb barriers, to return to the left toward the native state.[38] The figure shows only ensemble-averaged properties as a function of δ. There are variations among the conformations in

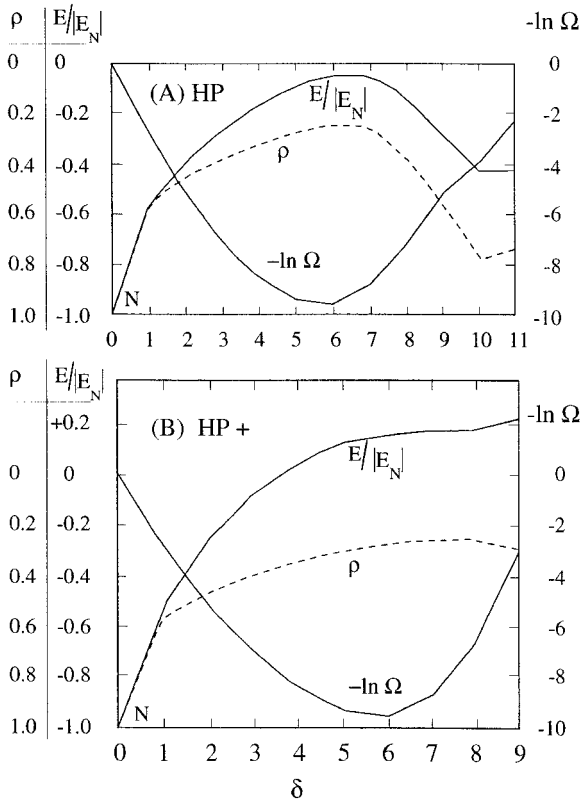

Fig. 4. Various averaged properties as a function of kinetic distance δ from the native state, for the sequence in Figure 1, in (**A**) the HP and (**B**) the HP+ energy functions. δ = 0 corresponds to the native state. Properties are averaged over all conformations of given δ. Average chain compactness is ρ (dashed curves, leftmost scale), defined as the number of intrachain contacts divided by the maximum possible number of contacts for the given chain length.[28] Average energy is *E* (solid curve, second scale from the left) is in units of the magnitude of the native energy, $|E_N|$. The number of conformations is Ω, −ln Ω is given by the scale on the right. The minimum in −ln Ω shows that most conformations are a distance δ ≈ 6 from the native state.

these ensembles, but the trend shown is representative.

Figure 4B shows that the HP+ model sequence folds along a smoother funnel than the same sequence in the HP model. Beginning folding at δ ≈ 4–7, the only way to reach lower energies, on average, is for conformations to slide downhill toward the native state at the left (see also Fig. 5). Nearly all the HP+ chains can reach the native state via fast throughway paths. The HP+ chain has some traps on its energy landscape (see Figs. 2, 6, and 7), but they are few and shallow, so their presence is not apparent in the averages in Figure 4B.

Figure 6 shows three example paths (I, II, and III) on the HP+ landscape. (We use the term "path" or "pathway" here to refer only to a single "microscopic" trajectory of a chain.) Figure 7 shows the energy changes along each trajectory. Figures 6 and 7 show that the landscape of the HP+ sequence is essentially a smooth funnel, insofar as there are few traps. (A more detailed discussion of the meaning of this "smoothness" is given below.) Most open conformations can cascade down to the native state without encountering a specific bottleneck conformation. The native structure can be approached by using many different microscopic trajectories.

Simulations such as ours lead to various kinds of information about the shapes of model energy landscapes. The properties shown in Figures 2 and 4 give rudimentary global properties that involve heavy averaging over the individual trajectories. This is like a distant aerial photograph of a mountain range that gives only a few highlights. At the other extreme, Figures 6 and 7 give the details of a few individual trajectories, as we would see in photographs of one or two individual walking trails on a mountainside. In principle, from information of the types shown in Figures 6 and 7, we could determine the full shape of the high-dimensional landscape, but this poses two problems: 1) the high dimensionality cannot be shown graphically, and what is worse, 2) as a process of multidimensional diffusion, any such figure is likely to be difficult to interpret, because it would be dominated by unimportant details.

The information of greatest value lies between these extremes of resolution. To try to convey the information from the simulations at a more meaningful level, we resort to the conceptual device of an artist's rendition of a landscape, such as is shown in Figure 5. These mock landscapes may be viewed as a projection of energy landscapes onto only three dimensions. Such diagrams keep only the main conceptual features. Figure 5a shows a bumpy landscape of the type followed by the HP sequence, and Figure 5b shows a funnelscape of the type followed by the HP+ sequence. Figure 5c shows a landscape with no throughway folding paths.

DYNAMICS OF FOLDING: CHAINS "POURING" INTO EVER LOWER ENERGY LEVELS

To obtain the ensemble-averaged time dependence of folding and unfolding, we use an analytical transition matrix method.[37,38] This can be done with traditional Metropolis sampling, but here we replace Metropolis sampling with Kawasaki dynamics[108] (see Appendix B). We do this because Metropolis sampling[85,109] has a physically unrealistic aspect for protein-folding kinetics. Metropolis sampling takes into account the slope of the landscape for an uphill step (say breaking a favorable contact), but it neglects the slope of the landscape for a downhill step (say making a favorable contact). This is adequate when Monte Carlo is used to compute the equilibrium properties of a system, as it was intended to do, but it is problematic for kinetics where the downhill slope of the energy landscape is important. Details of Kawasaki sampling are given in Appendix B.

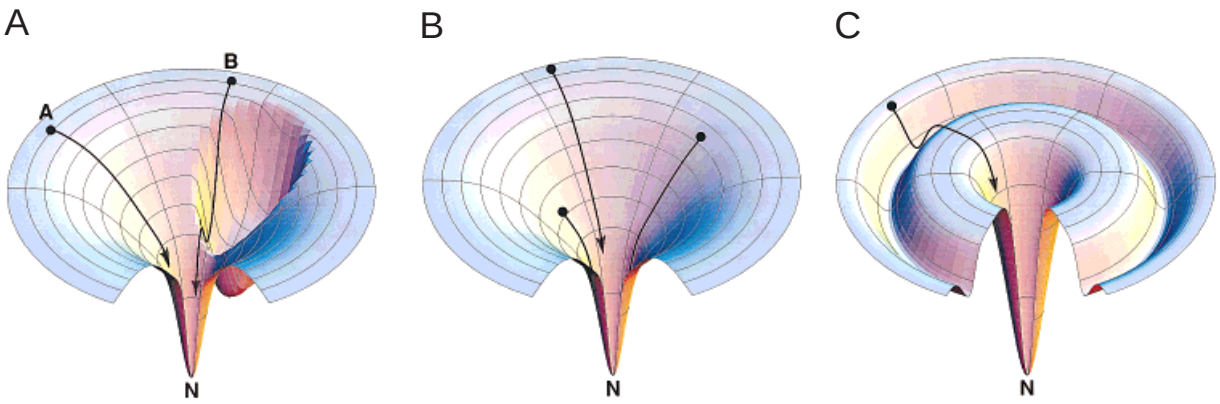

Fig. 5. General shapes of landscapes: **a:** The HP landscape is shown pictorially as having a kinetic trap, A is a throughway folding trajectory, whereas path B passes through the trap; **b:** The HP+ landscape is smooth, unfolding paths are simply the reverses of the folding paths shown (from Ref. 8). **a** and **b** are schematic representations of Figure 4. **c:** A landscape on which all folding molecules must pass through an obligatory folding intermediate, represented by the "moat" in the figure.

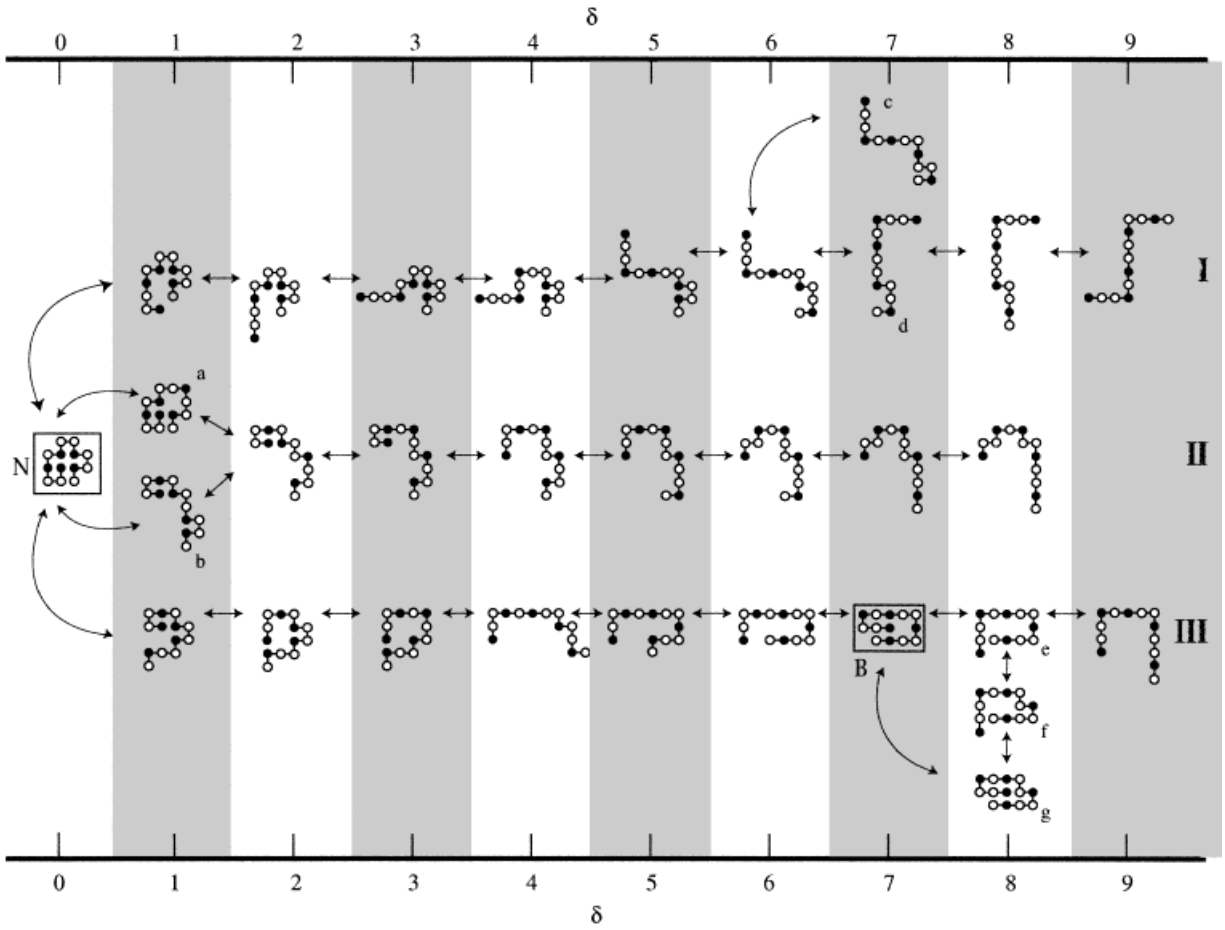

Fig. 6. Three individual folding trajectories in the HP+ model, as the folding reaction progresses to the native state ($\delta \rightarrow 0$). B is a kinetically trapped conformation that must break at least one favorable contact or make one unfavorable contact before it can reach native. No other conformations shown are trapped.

Figure 8 shows that the folding dynamics of the chain using the HP and HP+ energy functions are different. The HP chain has multiphasic folding. Approximately half the molecules fold to the native structure during the fast phase. The rest undergo slow folding. The slow folding phase in Figure 8

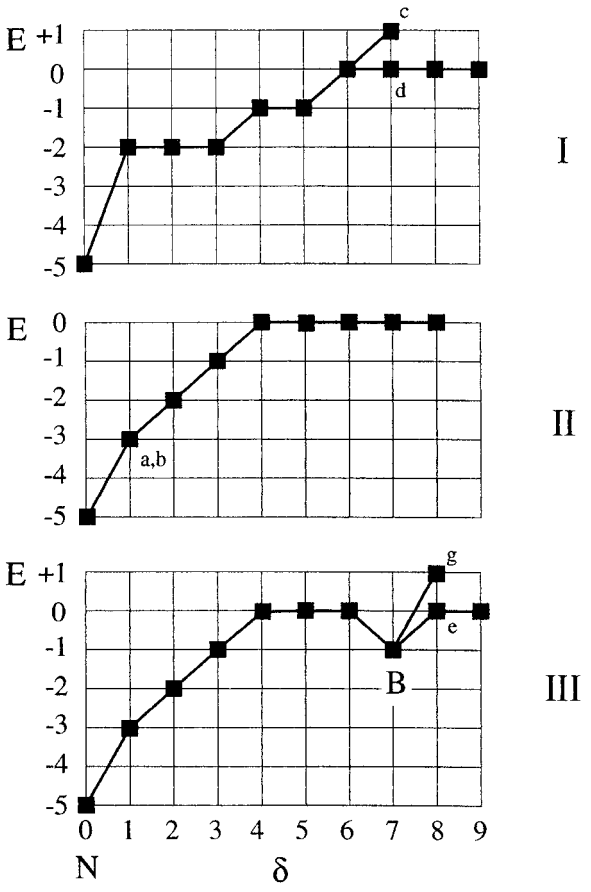

Fig. 7. Energy landscape trajectories in the HP+ model along paths I, II, and III in Figure 6. Energy E has units of $|\epsilon|$. Conformation labels a, b, c, d, e, and g correspond to those in Figure 6.

takes longer than the fast phase by two to three orders of magnitude. On the other hand, 90% of the HP+ chains fold in a single fast phase, about two orders of magnitude faster than the same sequence using the HP energy function.*

For the HP+ sequence, folding coincides with collapse (as measured by the energy in Figure 8). This is consistent with other model studies,[46,94,110,111] an interpretation of experiments on cytochrome *c* fast folding,[55,58] and recent experiments[112] on CspB. Figure 9 dissects the folding kinetics by energy levels. Chains "pour" from energy level 0 to −1 to −2, etc. with time.

Although Figure 9 appears "sequential," it differs from traditional sequential models of folding.[23,113,114] The difference is that the HP+ model shows very broad overlaps in time of ensembles that are themselves very broad distributions of conformations. For example, although the peak population of the −1 energy level occurs earlier in time than the peak of the −2 level, the populations of the two energy levels are otherwise nearly perfectly overlapping in time (see Fig. 9). Hence, these two populations cannot be described as separate "intermediate states" distinguishable from each other in time. These peaks are not equivalent to labels, such as I_1, I_2, etc., that define sequential intermediates that might be used

*The initial collapse of the HP sequence is faster than that of the HP+ sequence in Figure 8 because we used a stronger driving force for the former.

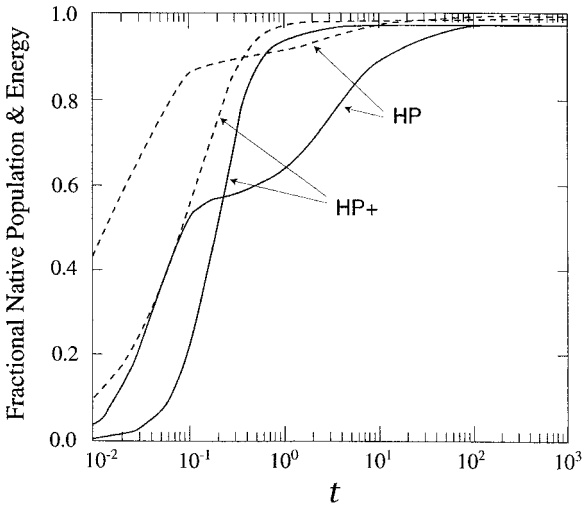

Fig. 8. Time dependence of folding in HP and HP+ models, computed by using reduced transition matrices and Kawasaki dynamics as described in Appendix B. Time evolutions are given at $\epsilon/(k_BT) = -6.16$ for the HP and $\epsilon/(k_BT) = -4.17$ for the HP+ sequence: these give equilibrium native populations equal to 98%. Solid curves show the native population; dashed curves show the average energy, in units of the native energy E_N, vs. model time t.

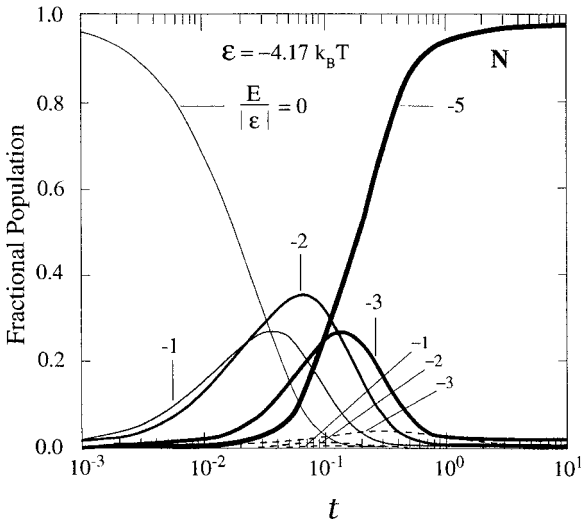

Fig. 9. Time dependence of subpopulations of conformations, for the sequence in Figure 1 in the HP+ model. Numbers indicate the energies. The native state N has $E/|\epsilon| = E_N/|\epsilon| = -5$. Trapped conformations are represented by dashed curves. Conformations with $E/|\epsilon| > 0$ are not shown because their populations are insignificant under folding conditions.

in mass-action models. More importantly, some conformations can "leapfrog" over some energy levels on their way downhill (see Fig. 2). Hence, the "pouring" we observe is more like the way water follows many different routes as it trickles down a complex hillside and less like the flow along a single downhill gulley.

UNFOLDING PROCESSES ARE NOT ALWAYS THE REVERSE OF FOLDING PROCESSES

Comparing Folding and Unfolding Dynamics in the Models

Mass-action models are usually based on comparing the unfolding and folding dynamics for a given protein. Which rate is faster? We monitor folding and unfolding relaxations in our models with the variable

$$\Delta y(t) \equiv \frac{[\mathrm{N}(t) - \mathrm{N}(\infty)]}{[\mathrm{N}(0) - \mathrm{N}(\infty)]} \qquad (1)$$

where N(t) is the fraction of the population that is native, as a function of the model time t defined in Appendix B. N(0) is the initial fractional native population at $t = 0$, and N(∞) is the equilibrium native population at $t \to \infty$. Folding starts from a subpopulation of the denatured state, with a uniform distribution among the $E = 0$ conformations. Unfolding starts with all chains in the native conformation. N(t) is computed by using the transition matrix method described in Appendix B.

Folding and unfolding of the HP and HP+ sequences are shown in Figures 10A and B, respectively. Relaxation of a second HP sequence (given in Fig. 11) is shown in Figure 10C. These plots are obtained with energies slightly favoring folding near the transition midpoint, similar to the conditions in some early protein relaxation experiments.[17,19,20]

The Three Classes of Folding Behavior Are Similar to Those Observed in Experiments

Figure 10 shows three different classes of behaviors found in the models: relaxations in the folding and unfolding directions happen at the same rate, or one rate is faster than the other. After a model deadtime, for $t > 0.2$–0.4, Figure 10A shows equilibration is faster for folding than for unfolding for the first HP sequence. Figure 10B shows a sequence for which folding and unfolding equilibration rates are about the same, and Figure 10C shows an HP sequence for which unfolding equilibration is faster than folding. These behaviors are readily understood in terms of their energy landscapes (see Figs. 2 and 11). Upon jumping conditions from native stabilizing to the denaturation midpoint, the main route of unfolding is through the lowest-energy conformations, i.e., small steps up in energy from the native state. Hence, unfolding kinetics is sensitive to the structure of the low-energy part of the landscape near the native state. For such cases, the dominant unfolding paths are indicated by arrows in Figures 2 and 11.

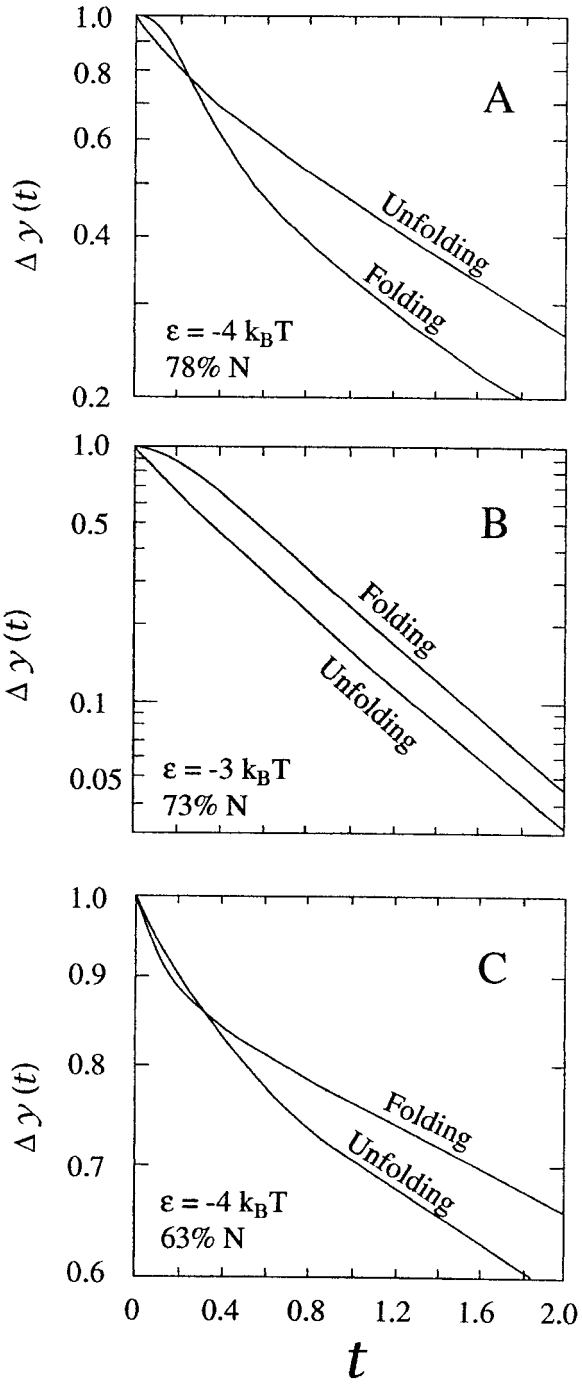

Fig. 10. Relaxation decay curves from model simulations for (**A**) the HP model, (**B**) the HP+ model, and (**C**) the sequence in Figure 11, in the HP model, at the given intrachain interaction strengths ϵ, which correspond to the equilibrium N populations shown.

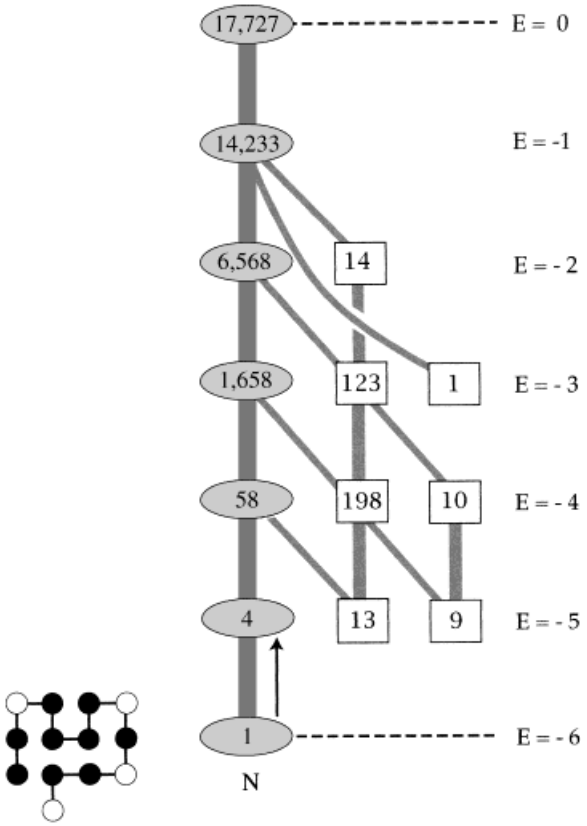

Fig. 11. Conformational flow diagram for the second sequence, using the HP energy function. The labeling scheme is the same as in Figure 2B.

On the other hand, Figure 10B shows that on "funnelscapes," unfolding is nearly a perfect reversal of the folding process. Kinetic traps have little effect on the folding and unfolding of the HP+ chain (Fig. 10B). Thus, interconversions between the native and denatured conformations are fast on funnelscapes, because they are essentially unimpeded by energy barriers. The only difference occurs at the very early stage of folding, during which there is a very fast ($t < 0.4$) pouring of denatured conformations from the initial $E = 0$ ensemble into lower energy states (see Figs. 5b and 9).

When the unfolding relaxation rate differs from the folding relaxation rate (in Figs. 10A,C), it is due to the distribution of kinetic traps on the model energy landscapes (Fig. 12). Consider the HP chain in Figure 10A. Why is folding relaxation faster than unfolding? Figure 8 shows that half of these chains fold quickly, without surmounting energy barriers. But unfolding is slower because *all* the chains leaving the native state must surmount one particular energy barrier to reach the most stable denatured conformations. For $\epsilon = -4k_BT$, the main equilibrium denatured populations are the $d = 1$ and $d = 2$ states with $E = -4$ (45% and 13%, respectively). Figure 2A shows that there is no direct kinetic connection from the native state to these stable denatured states. Rather, native chains must first pass through one of the 257 $E = -3$ (high-energy) conformations before they can drop back down to the $E = -4$ trapped conformations that dominate the denatured ensemble under these conditions (Fig. 12b). Figure 13 shows one such unfolding path. In such cases, all unfolding routes have an energy barrier with a common height, whereas some of the folding routes have no energy barrier.

Figure 10C shows the opposite situation: folding relaxation is slower than unfolding. For this chain sequence, there are fewer throughway folding routes; most folding chains get trapped (see Fig. 11). Early unfolding relaxation is faster than folding because four of the most stable denatured conformations (13% of the equilibrium denatured population) are along throughway unfolding paths; they are directly accessible from the native state (Fig. 12c).

The Bumps on Landscapes: the "Ups" Are Not the Same as the "Downs"

The speed of folding is determined by more than just a simple average of the bumpiness of an energy landscape. Some bumps are basins (traps) and others are hills (see Fig. 14). Hills are energetically unfavorable conformations, caused by repulsive interactions or conformations with van der Waals overlaps, for example. Hills are less troublesome obstacles than traps are to folding. Aside from Brownian buffetting, on average, a folding chain will proceed downhill in energy. Where there are basins, the chain will get caught in them, but where there are hills, it will go around them. Energy levels alone are insufficient to account for the kinetic properties of an energy landscape. Kinetic connectivities among the different conformational states are also important.[4,37,38] For example, the distribution of energy levels of the denatured state is much wider for the HP+ than for the HP sequence (Fig. 2). The standard deviation about the mean energy of the denatured ensemble is $1.34|\epsilon|$ for the HP+ sequence and $0.66|\epsilon|$ for the HP sequence, implying that the HP+ landscape is "bumpier" by this gross measure even though chains fold much faster on it than on its HP counterpart. The fact that the *trajectories* of HP+ folding look mostly like smooth funnels (it is in this sense that the HP+ landscape is "smooth", see Figs. 6 and 7) hides the fact that the *full landscape* has many hills. Some spin-glass models characterize landscapes by their "average bumpiness" as measured by the standard deviations of energies.[25,26] But not all bumps are created equal!

It has been suggested that the existence of a pronounced "energy gap" between the native ground state and the next state higher in energy (first

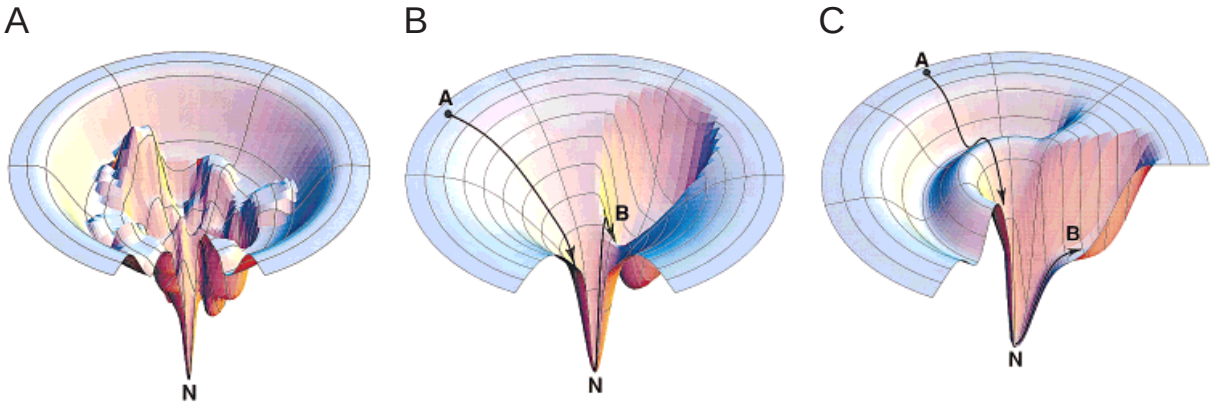

Fig. 12. Different folding scenarios. The vertical axis is internal free energy. Each conformation is represented as a point on the landscape. The two horizontal axes represent the many chain degrees of freedom. **a:** A rugged landscape with hills and traps, folding kinetics is likely multiple-exponential (from Ref. 8). **b:** A landscape in which folding is faster than unfolding. A is a through-way folding path, whereas unfolding chains (path B) must surmount a barrier to reach the most stable denatured conformations. **c:** A landscape in which folding is slower than unfolding. Most folding paths (path A) pass through a kinetic trap, whereas some low-lying denatured conformations are readily accessible from the native state during unfolding (path B).

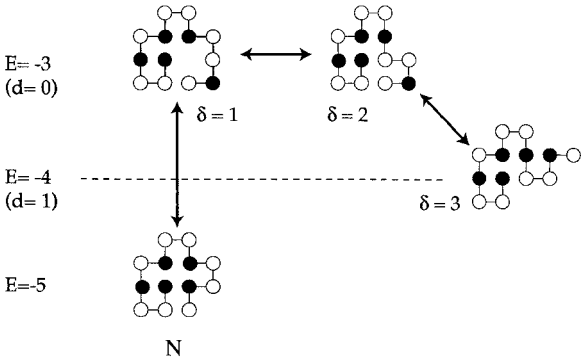

Fig. 13. Climbing over an obligatory barrier in *unfolding*, from the native (**N**) conformation ($E = -5$) to a lowest-energy nonnative conformation ($E = -4$), for the HP sequence in Figure 2A. The figures shows one of many such trajectories.

excited state) is the "necessary and sufficient" condition for fast folding in certain heteropolymer models[5,100,101] (see also Ref. 102). By standard mathematical terminology, necessary and sufficient means "no exceptions." Such a statement is too strong to account for our results. Nevertheless, we do observe some correlation between folding rate and energy gap. Our faster folding HP+ sequence has a larger energy gap than the two slower folding HP sequences here (2 units vs. 1 unit). However, folding rates are also dependent on the kinetic connectivities among conformations[37,38] as well as on temperature.[39,40,96,115] The two HP sequences in Figures 10A and C have identical energy gaps, but the first HP sequence (Fig. 10A) folds approximately twice as fast as the second HP sequence (Fig. 10C), for the same intrachain interaction $\epsilon = -4k_BT$. Lack of correlation between folding rate and the size of the energy gap has also been shown in recent model studies.[96,110,111,116]

Are Experimentally Observed "Folding Intermediates" Due to Kinetic Traps?

Folding experiments show all three possible behaviors: folding and unfolding relaxation rates are equal, or one is faster than the other. Corresponding to Figure 10B, where folding and unfolding proceed at identical rates are the experiments of Ikai and Tanford[17] on lysozyme in guanidine. This follows the two-state model N ⇌ D (N is the native and D is the denatured state).

In other systems, equilibration rates of fast and slow phases can differ by many orders of magnitude.[17,22,24] Ikai and Tanford[17] found for cytochrome *c* that early-stage relaxation is faster for folding than for unfolding (see also Refs. 19 and 20), corresponding to Figure 10A. On the other hand, Tsong et al.[21,22] observed that for ribonuclease A the initial relaxation for unfolding is faster than for folding.

These data have been interpreted in terms of "folding intermediates," which are represented as symbols inserted into mass-action kinetics laws. Ikai and Tanford[17,18] found that their cytochrome *c* relaxations did not fit well the so-called "on-pathway" model

$$\mathrm{N} \rightleftharpoons \mathrm{X} \rightleftharpoons \mathrm{U} \qquad (2)$$

where N, U, and X are the native, fully unfolded and intermediate states. They found that cytochrome *c* is better described by either an "off-pathway" model

$$\mathrm{N} \rightleftharpoons \mathrm{U} \rightleftharpoons \mathrm{X} \qquad (3)$$

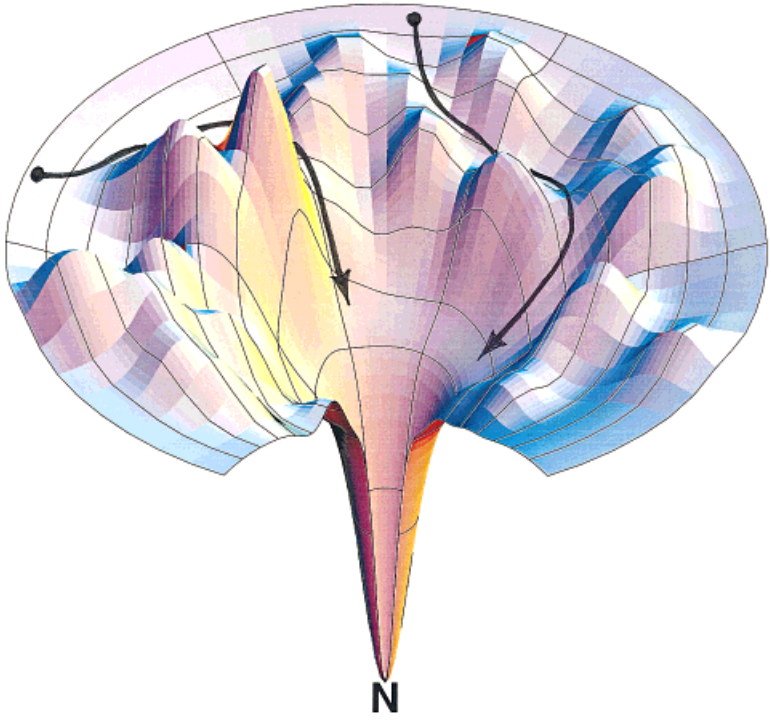

Fig. 14. Funnelscape for a fast-folding protein. Folding is limited by the rate of meandering downhill.

or a four-state model

$$N \rightleftharpoons X_1 \rightleftharpoons U \rightleftharpoons X_2 \qquad (4)$$

where X_1 and X_2 are, respectively, "on-pathway" and "off-pathway" intermediate states. X states are intermediate in the sense that they have average optical properties that are between those of N and U. Ikai et al.[17,19] recognized that intermediates may have "close contacts between residues that are not in close contact in the native structure, and they must clearly be reversed before the native structure can be attained." This view of cytochrome *c* kinetics is consistent with Figure 10A, where traps slow the folding process. Recent experiments support the view that nonnative contacts in folding intermediates decelerate folding of cytochrome *c* under certain solution conditions.[55,58]

The data on ribonuclease A (Tsong et al.[21,22]) has been interpreted instead in terms of a sequential model[23]

$$U \rightleftharpoons I_1 \rightleftharpoons I_2 \rightleftharpoons \cdots I_n \rightleftharpoons N \qquad (5)$$

where U and N are the fully unfolded and native states, respectively, and the I_1, I_2, . . . I_n have been called on-pathway intermediates. In addition to the two main kinetic phases observed by Ikai et al.[19] in guanidine-induced unfolding, Tsong[24] later observed a fast third kinetic phase in thermal unfolding and folding of ferricytochrome *c* at acidic pH, suggesting that the sequential model might also apply to cytochrome *c* under some conditions. In the sequential model, the rapidity of initial unfolding was explained as involving only one step from N to I_n, whereas folding is slower because it requires many steps through intermediates.

Kinetic Traps Can Account for Both "On-Pathway" and "Off-Pathway Intermediates"

Although the ribonuclease data and Figure 10C show qualitatively the same behaviors, our interpretation of model results in Figure 10C is different from that of Tsong et al.[21,22] Rather than "on-pathway intermediates," the slowing of folding in Figure 10C is caused by traps in the energy landscape (see Fig. 11), just as it is for Figure 10A. Hence, what is explained in the mass-action models as "on-pathway" vs. "off-pathway" intermediates, in our models is just a difference in the distribution of traps on the energy landscape.

Strictly speaking, structured intermediates *must be in energy traps;* otherwise, they would not accumulate and be observed as intermediates. Any stable or metastable structure on an energy landscape must involve some sort of basin. (Here, energy means the internal free energy of a conformation—the vertical scale on the energy landscape figures.) Such intermediates correspond to "misfolds"[2,55,58,117,118] resulting from formation of low-energy nonnative contacts or steric hindrance, either of which can stall further progress.[38]

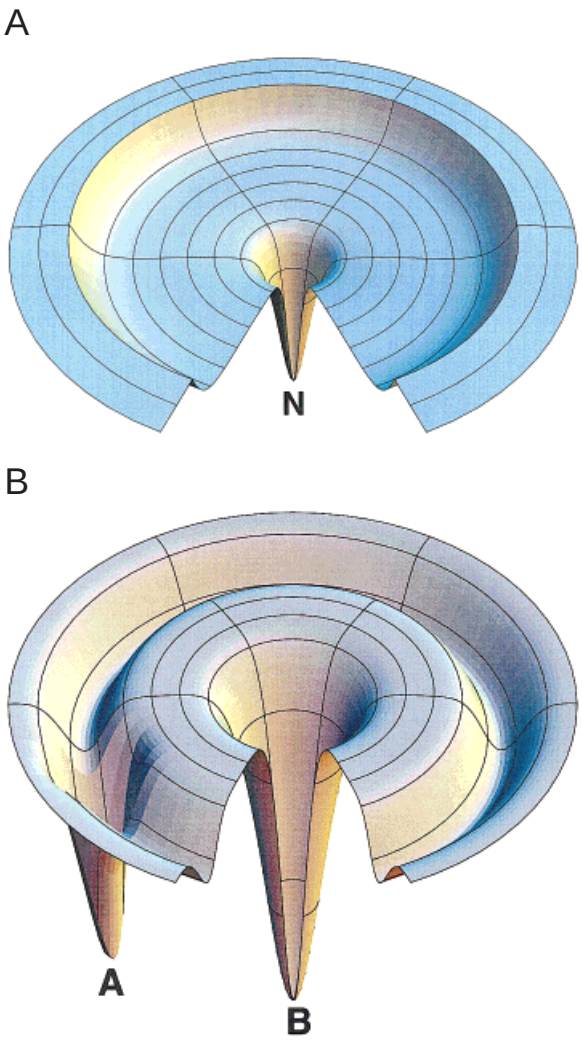

Fig. 15. **a:** Champagne glass landscape, to illustrate how conformational entropy can cause "free-energy barriers" to folding. The "bottleneck" or rate limit to folding is the aimless wandering on the flat plateau as the chain tries to find its way downhill (from Ref. 8). **b:** Serpin scenario shows a landscape with a deep kinetic trap on the left (A), which is easily accessible from the open conformations. Chains trapped in this deep local minima anneal to the global minimum (B, in the middle) only very slowly. This may correspond to the folding of some serpins, such as PAI-1.

Kinetic Intermediates as Broad Conformational Ensembles

But there is also a second type of "stable intermediate" in the energy landscape perspective. Folding can be slowed by conformational entropy barriers, which are more like *plains* on energy landscapes, not *basins*. Free-energy "barriers" occur when folding is slow due to extensive conformational searching, where the energy landscape may be relatively flat (Fig. 15a; see also Refs. 119–121). In this case, there is no one specific trapped conformation. In fact, it is the large size of the ensemble that makes searching such a slow process in those cases. Although there *is* a barrier in the *total free energy* in these cases, there is no *barrier* on the *energy landscape*, because the vertical axis is the internal free energy (all energies except for the conformational entropy). This scenario may describe some intermediate states that have diverse and significantly expanded conformations.

In Some Landscapes the Global Minimum Is Kinetically Inaccessible From the Denatured Ensemble

Figure 15b shows an energy landscape that may describe the observed folding kinetics of some serpin proteins, such as plasminogen activator inhibitor-1 (PAI-1).[122–125] PAI-1 folds rapidly to a unique conformation that is biologically active but is metastable (represented by the local minimum A in Fig. 15b). Over a longer timescale, this conformation converts to a "latent" conformation (B) that is nonfunctional but is thermodynamically more stable against denaturation.[72,122] Therefore, even if the more stable latent form is the global minimum conformation of PAI-1, it is inaccessible from the ensemble of denatured conformations during the time in which PAI-1 is functional.

In the absence of its pro region, α-lytic protease may be a protein with a similar energy landscape. Similar to the PAI-1 scenario, it folds rapidly to a molten globule-like intermediate state and much more slowly to its native structure.[126] However, the intermediate state of α-lytic protease may not be unique, because it is significantly expanded relative to its native state. Hence, the single deep well in Figure 15b is likely absent in its energy landscape. Instead, it may resemble more the moat landscape in Figure 5c or the champagne glass landscape in Figure 15a with an entropic barrier.

Enthalpic Bumps on Energy Landscapes

The landscapes in this paper are shown coarse grained. They do not show the many small bumps that are the relatively low enthalpic barriers against bond rotations (see Appendix C). Thus, even the conformational search-limited processes described above probably have some enthalpic components, due to surmounting the many small rotational barriers that separate energy minima of dihedral angles. We believe that these are likely to be the enthalpic origins of free-energy barriers on funnel landscapes (see, for example, Refs. 54, 70, 112).

Experimental Characterization of Folding Intermediates

What are experimental folding intermediates? Among the best understood examples are those due to proline isomerization.[2,113,127–131] Proline isomerization has been invoked to account for cytochrome *c* folding.[132] For example, double-jump experiments by Ridge et al.[133] suggest that the slow phase in ferricytochrome *c* folding observed by Ikai et al.[19] is consis-

tent with proline isomerization. Proline isomerization has been incorporated into the sequential model and applied with some success to account for the accumulation of intermediates in kinetic studies of ribonuclease A (Refs. 114 and 134).

But proline isomerization does not appear to account for all slow kinetic phases in protein folding and unfolding (reviewed by Kim and Baldwin[130]). For instance, the slow refolding kinetic phase of human lysozyme does not result from proline isomerization.[135] Proline isomerization decelerates folding in ribonuclease T1, but it also involves a specific contact and steric hindrance.[136] Nor is proline isomerization solely responsible for the folding behavior of cytochrome *c* (Refs. 55, 58; T.R. Sosnick, private communication; see also Ref. 137) or *trp* aporepressor of *Escherichia coli*.[138]

Hence, multiple exponentials in folding dynamics are not always explained by proline isomerations. Multiple exponentials observed in experiments may also be due to traps on energy landscapes. Equilibrium molten globules of α-lactalbumin[139–142] and lysozyme[139,140] are found to be close analogs of their early kinetic folding intermediates. These findings appear consistent with a sequential "molten globule" model of folding (reviewed in Refs. 2, 143, and 144). But other proteins appear to fold by multiple-path, nonsequential mechanisms, and folding intermediates may correspond to kinetic traps (reviewed in Refs. 1, 2, and 145). The predominantly β-sheet protein β-lactoglobulin passes through a helical state upon folding.[146–151] This intermediate cannot be "on-pathway," at least insofar as "pathway" implies a monotonic decrease of the rms deviation of the chain from the native structure, for example.

There is additional experimental evidence that some folding does not follow a sequential mechanism. The folding of hen egg white lysozyme under strong folding conditions appears to follow a nonsequential mechanism.[117] Folding consists of a fast hydrophobic collapse, followed by slow formation of fixed tertiary interactions.[60,117,118,152] There are parallel alternative folding pathways.[60,117] The α-helical domain folds faster than the β-sheet domain. There is an overshoot of the circular dichroism (CD) signal similar to that observed for β-lactoglobulin. It has been suggested that this overshoot may not correlate with secondary structure.[153] However, it is also possible that some folding intermediates of lysozyme are kinetically trapped conformations, which may then require some conformational opening before the proteins can proceed to the native state.[35,37,38,117] Also, in apparent disagreement with the simple sequential picture of folding,[134] pulsed hydrogen exchange experiments show that folding kinetics of different parts of the protein are not synchronized (see, for example, Ref. 154 for oxidized cytochrome *c;* Ref. 155 for barnase; reviewed in Ref. 2 by Baldwin).

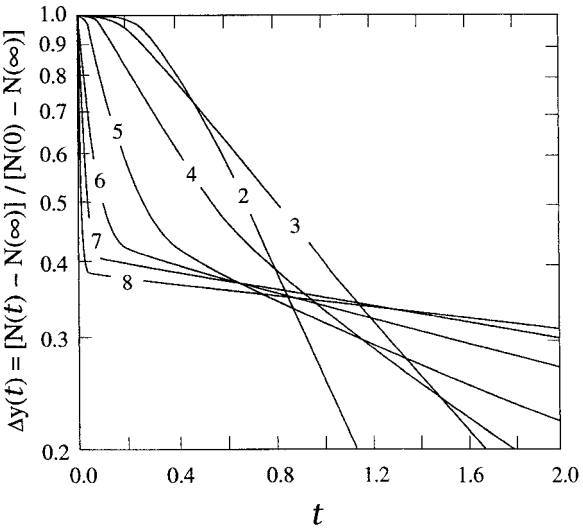

Fig. 16. Relaxation decays for folding, for the HP sequence (Fig. 1) under different driving conditions (different HH sticking strengths). The labels are $-\epsilon/(k_B T)$.

SOME LATTICE MODEL PROTEINS HAVE CHEVRON PLOT KINETICS

HP Sequences Often Have Bumpy Landscapes With Traps

Figure 16 shows essentially single-exponential, two-state folding under weak folding conditions (except for an initial fast adjustment for $t < 0.4$), and multiple-exponential folding under stronger conditions. Approximately 50–60% of the chain population can fold via throughway paths (Fig. 8). Under conditions that only weakly favor the native structure, the traps are shallow, making escape easy. More strongly native conditions help stabilize both the native state and the kinetic trap conformations, so traps then play a more prominent role in the folding dynamics. A similar crossover between single-exponential and multiphasic folding has been observed in hen lysozyme.[17,20,117] This result is also consistent with three-dimensional lattice Monte Carlo simulations of first passage time to the native state at different temperatures.[40]

Figure 17 shows that the lattice model gives chevron behavior of the relaxation rate as a function of denaturant concentration[63,64,66–68] for the fast-phase relaxation. Figure 16 shows that within a certain range, $0.5 < \Delta y < 0.9$ for example, folding relaxation is approximately exponential

$$\Delta y(t) = y_0 \exp(-k_r t) \tag{6}$$

where k_r is an apparent relaxation rate. Figure 17 gives the folding relaxation rate† k_r for $0.5 < \Delta y <$

†In the present model, y_0 is not exactly unity and is slightly different for different ϵ's because of the initial ($t < 0.4$) slow accumulation of native population under weakly native conditions.

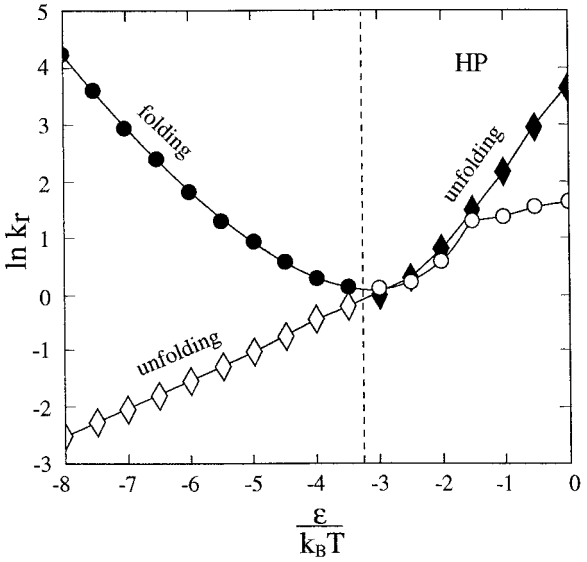

Fig. 17. Model chevron plots for the HP sequence. Relaxation rates ln k_r vs. ϵ estimated between $\Delta y \approx 0.75$ and $\Delta y \approx 0.5$, for folding (circles) and unfolding (diamonds) of the HP sequence in Figure 16. Filled symbols represent folding relaxation under folding conditions and unfolding relaxation under unfolding conditions. The dashed line indicates the equilibrium folding-denaturation midpoint.

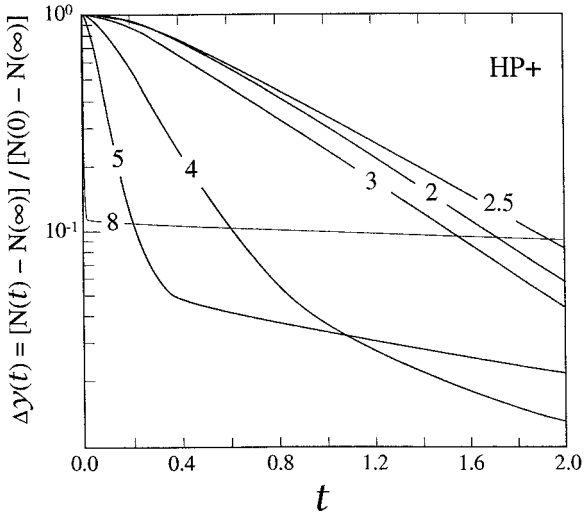

Fig. 18. Same as Figure 16, but for the HP+ sequence.

0.75 as a function of $\epsilon/(k_BT)$, the quantity that mimics the denaturant strength.[156,157] As in most experiments, Figure 17 shows that the folding relaxation under folding conditions and unfolding relaxation under unfolding conditions combine to form a typical chevron plot. Figure 17 also shows that unfolding relaxation slows monotonically as HH sticking increases (ϵ becomes more negative). The cause of this behavior is the same as for Figure 10A.

HP+ Sequences Often Have Smooth Funnel Landscapes

Figures 18 and 19 show that the HP+ chain has a much smoother landscape than the HP chain, in the sense that effects of kinetic traps are not significant (see above). The fast kinetic phase accounts for approximately 90% of the chain population. For a given HH sticking energy, folding and unfolding relaxation rates are approximately equal (see also Fig. 10B). In contrast to the HP sequence results in Figure 17, *both* the folding and unfolding kinetic data of the HP+ sequence show chevron behavior.[67] In this regard, the simulation results of Figure 19 resemble experiments on chymotrypsin inhibitor 2 (CI2),[53,54] cytochrome *c* (Refs. 55 and 58), monomeric λ repressor,[56,57] and CspB (Refs. 59 and 61).

Nonchevron Behavior in Multistate Kinetics

Figure 20 shows that although the fast phases of HP and HP+ chains follow chevron behavior, the slow phases do not. For both HP and HP+, the slow relaxation rates are practically identical for folding and unfolding (individual relaxation curves not shown). Slow-folding phases involve barrier climbing. The more native-like the conditions, the deeper are the traps and the higher are the relative barriers, so the slower is the relaxation. Hence, decreasing denaturant (making ϵ more negative) results in monotonically slower relaxation, as is observed experimentally[136,142,158] and in other models.[159]

Similar trends of decreasing slow-phase relaxation rates with decreasing denaturant concentration are observed in experiments on cytochrome *c*,[19,58] and lysozyme,[20] carbonic anhydrase B,[160] ribonuclease T1,[136] a mutant of dihydrofolate reductase,[158,161] mutants of staphylococcal nuclease,[162] and some fragments of the α subunit of tryptophan synthase (J.A. Zitzewitz, private communication). The rate of formation of the native disulfide bond in the constant fragment of the immunoglobulin light chain[163] and the rate of rearrangement of a nonnative disulfide folding intermediate of and BPTI[164] also decreases with decreasing urea concentration.

Figure 21 shows that for proteins having complex landscapes, sometimes the relaxation rate is not a linear function of denaturant concentration. The unfolding mechanism can change with denaturant. Under weakly native conditions, the native state of the HP sequence shown in Figure 11 is kinetically adjacent to many different conformations in energy levels $E = -5$, $E = -4$, $E = -3$, and $E = -2$, so the chain denatures by passing through all of them. But under strongly native conditions, the chain only passes through the lowest-energy denatured conformations, so the routes of unfolding are then more restricted. Therefore, k_u varies nonlinearly with ϵ. In the present model, exact kinetic adjacency data of this sequence and the results of Appendix B lead to the exact unfolding rate $k_u = \exp[\epsilon/(2k_BT)] + 5\exp[\epsilon/(k_BT)] + 3\exp(3\epsilon/(2k_BT)] + 4\exp[2\epsilon/(k_BT)]$ for this sequence. This rate is plotted in Figure 21.

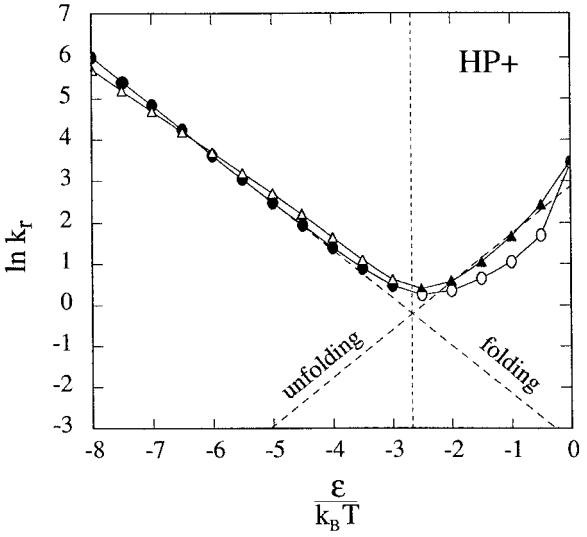

Fig. 19. Model chevron plots for the HP+ sequence. Same as Figure 17, but for the HP+ sequence. The inclined dotted lines are linear approximations to the folding and unfolding rates. Here folding and unfolding relaxation rates are represented, respectively, by ovals and triangles (from Ref. 8).

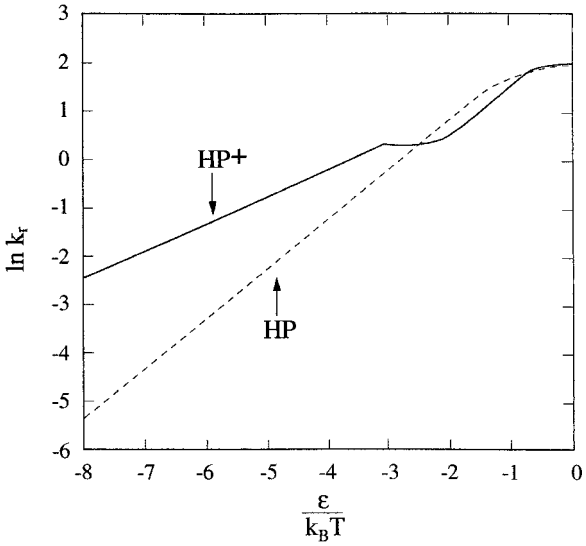

Fig. 20. Slow kinetic phases show nonchevron behavior for the HP and HP+ sequences. Relaxation rates ln k_r vs. ϵ estimated between $\Delta y = 10^{-4}$ and $\Delta y = 10^{-6}$ for folding and unfolding of the sequence in Figure 1. For each sequence, folding and unfolding relaxation rates are practically indistinguishable.

Linear extrapolation over wide ranges of denaturant concentrations are often used for equilibrium (Refs. 165, 166, and references therein) and kinetic[71,73] studies of proteins. B.-L. Chen et al.[71] pointed out that long extrapolations of kinetic rates are not valid in some cases. Our model results in Figure 21 also indicate that such procedures can be problematic for some energy landscapes.

THE NON-ARRHENIUS TEMPERATURE DEPENDENCE OF FOLDING AND UNFOLDING

The bottlenecks, transition states, and activation barriers for folding are typically probed by measuring the temperature dependence of folding and unfolding kinetics. Here, we find that the lattice models predict non-Arrhenius behaviors that closely resemble experiments. We focus on the HP+ sequences having smooth funnelscapes. As is standard for experiments,[70] we make linear approximations to the folding and the unfolding rates by extrapolating relaxation rates from strongly folding (k_f) and strongly denaturing (k_u) conditions:

$$\ln k_f \equiv \left.\frac{dN}{dt}\right|_{N=0} \approx -3.41 - 1.18\,\frac{\epsilon}{k_B T}$$

$$\ln k_u \equiv -\left.\frac{dN}{dt}\right|_{N=1} \approx 2.80 + 1.15\,\frac{\epsilon}{k_B T}. \qquad (7)$$

As in the conventional two-state picture,[70,71] the linearized folding and unfolding rates in Eq. (7) are identical (the two dashed lines in Fig. 19 intersect) at the denaturation midpoint $\epsilon = -2.61 k_B T$, and the

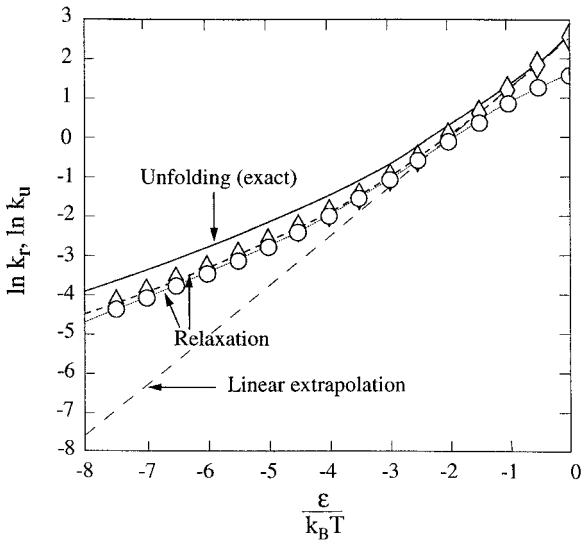

Fig. 21. Nonchevron behavior. Relaxation rates ln k_r vs. ϵ estimated between $\Delta y \approx 0.6$ and $\Delta y \approx 0.1$, for folding (circles) and unfolding (diamonds) for the sequence in Figure 11. The solid curve corresponds to the exact unfolding rate (ln k_u). The dotted line indicates the linear extrapolation of the unfolding rate from the relaxation rate under strongly denaturing conditions.

total relaxation rate is approximately given by

$$k_r \approx k_f + k_u \qquad (8)$$

for both folding and unfolding relaxations.

We use the linearized approximations in Eq. (7) to express the equilibrium free energy of folding as

$$\frac{\Delta G_{fold}}{k_B T} \equiv -\ln \frac{N}{1-N} \approx \ln \frac{k_f}{k_u} \approx 6.21 + 2.33\,\frac{\epsilon}{k_B T}. \qquad (9)$$

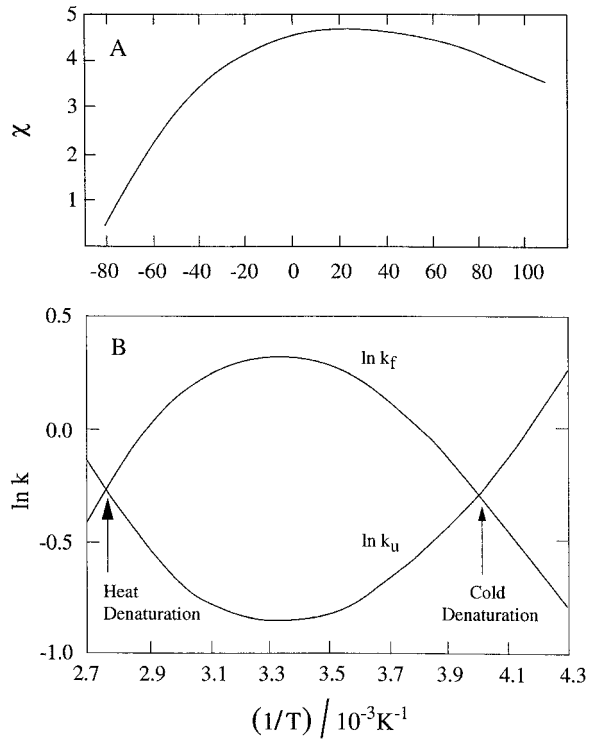

Fig. 22. **A:** Temperature dependence of the hydrophobic interaction, χ, parameterized from experimental data (from Fig. 2 of Ref. 167). Folding rate k_f and unfolding rate k_u are approximated by linear functions of $\epsilon/(k_BT)$. **B:** Predicted temperature dependence trend of folding and unfolding rates, k_f and k_u for $\epsilon(T)/(k_BT) = -C\chi(T)$, where $\chi(T)$ is given by (**A**). We used $C = 0.67$ to illustrate the possibility of both heat and cold denaturation.

This is a reasonable approximation for $-5 \leq \epsilon/(k_BT) \leq -1.5$.

To study the temperature dependence of folding requires knowledge of the temperature dependence of the hydrophobic interaction $\epsilon(T)$. We use an empirical expression obtained earlier[167] based on oil/water partitioning experiments and plotted in Figure 22A. So far in the present paper, we have studied folding and unfolding rates as a function of $\epsilon/(k_BT)$. Now, by inserting $\epsilon(T)$ from Fig. 22A, we can describe the folding and unfolding rates vs. temperature. Figure 22B shows that this model gives qualitative agreement with experiments on the temperature dependence of protein folding kinetics.[54,58,61,65,70,73,76,77,168] Our model results and experimental data share the common features that the ln k vs. $1/T$ plot for folding is concave downward, whereas for unfolding it is concave upward. These curvatures are similar to those of the cold denaturation experiments of Chen et al.[70] (see also Appendix C and Ref. 8).

Non-Arrhenius Behavior: Due to Kinetic Traps or the Hydrophobic Effect?

Two explanations have been proposed for the non-Arrhenius temperature dependence of folding kinetics. One view is that the curvature is due to kinetic traps on the energy landscape. That is, weak native conditions cause slow folding because there is little driving force. Then, more strongly native conditions cause faster folding. Even more strongly folding conditions deepen the kinetic traps and again slow the folding, giving rise to folding rates vs. $-\epsilon/(k_BT)$ that are concave downward.[33,38,39,42–44,48] This behavior is quite typical of sequences that have multiphasic kinetics and was first observed in the HP model[33,38] (reviewed in Refs. 3 and 4; see also Fig. 31A in Appendix B). On that basis, it has been suggested by Bryngelson et al.[3] and Galzitskaya and Finkelstein[44] that kinetic traps may account for the non-Arrhenius $1/T$-dependences of folding rates.

However, our results above suggest that most of the curvature in the plot of ln k vs. $1/T$ for *fast-folding, two-state* proteins with single-exponential kinetics (modeled by the HP+ sequence) is mainly due to the temperature dependence of the hydrophobic interaction. Rates in HP+ kinetics depend on the quantity $\epsilon/(k_BT)$ that models the Flory χ parameter. In the HP+ model, the linear dependence of rates on denaturant concentration (chevron plots) and their non-Arrhenius dependence on $1/T$ arise from the linear dependence of $\epsilon/(k_BT) \propto \chi(c, T)$ on denaturant concentration c at a given T and the nonlinear dependence of $\chi(c, T)$ on $1/T$.[156,167] This picture is consistent with recent experiments on Protein L and analyses of CspB data by Scalley and Baker[169] (see also Appendix C). In fact, if the non-Arrhenius behavior were caused by traps, the changes in barrier heights caused by variations in $\epsilon/(k_BT)$ should give a nonlinear variation of ln k_r along the folding branch (where ln $k_r \approx$ ln k_f), rather than the experimentally observed linear chevron behavior as a function of denaturant concentration. Moreover, models with T-independent contact interactions cannot account for cold denaturation. For proteins with two-state kinetics, the curvatures in the logarithm of unfolding rates (ln k_u) vs. $1/T$ and the linear relation between ln k_u and denaturant concentration simply cannot be both arising from kinetic traps, but these two behaviors have been observed concomitantly in protein kinetic experiments,[61,69,70,77] including proteins that undergo cold denaturation.[69,70]

CHEVRON PLOT "ROLLOVER" AND INTERACTION SPECIFICITY

On the other hand, we believe that the "rollover" behavior[170] found in some chevron plots is likely to result from kinetic traps. Rollover is where chevron plots flatten out at very low denaturant, or even where folding rates decrease as denaturant concentration decreases to zero. Hen lysozyme,[60] wild-type and a mutant of barnase,[171] and ubiquitin variants[172] deviate from perfect chevron behavior under strongly folding conditions. Cytochrome *c* also exhibits rollover behavior.[58] In pH-induced folding or

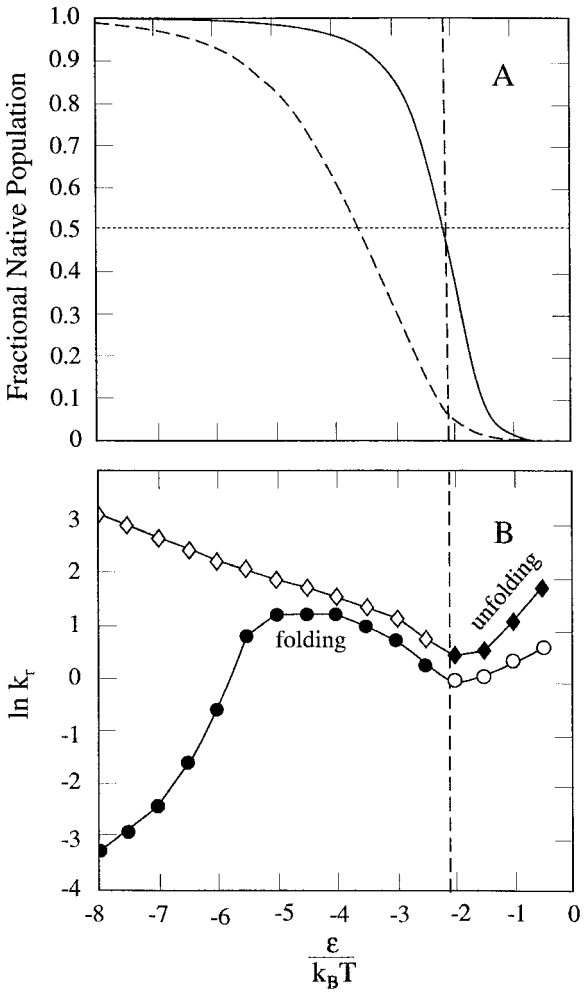

Fig. 23. Augmented model that also includes solvent-independent energy penalty $p_{nn} = 3k_BT$ for every nonnative contact. **A**: Equilibrium denaturation curve (solid curve), compared with the unmodified HP model ($p_{nn} = 0$, dashed curve). The vertical dashed line marks the transition midpoint. **B:** Relaxation rates ln k_r vs. ϵ, estimated between $\Delta y \approx 0.75$ to $\Delta y \approx 0.5$, for folding (circles) and unfolding (diamonds). Filled symbols indicate folding relaxation under folding conditions and unfolding relaxation under denaturing conditions.

unfolding of staphylococcal nuclease, the leveling off of chevron plots under strongly folding conditions has been observed.[173]

A physical cause of rollover can be an interaction that is unaffected by denaturant. Figure 23 shows the lattice model results augmented by an additional energy of this type (ϵ-independent penalty $p_{nn} = 3k_BT$ for each nonnative contact) for the HP sequence in Figure 11. With this modified energy, the sequence is more stable (Fig. 23A) because the energy spectrum of the denatured state is shifted up by the repulsion. In contrast to the nonchevron behavior of Figure 21, unfolding relaxation in Figure 23B is chevron-like. Folding relaxation in Figure 23B is partly chevron-like, but the rate of folding relaxation decreases with increasing HH sticking under strongly folding conditions.

This model gives chevron-like behavior under weakly native conditions because there are few traps when the repulsive interaction p_{nn} is relatively strong. But as conditions become more strongly native, the repulsive interactions are overcome by the HH sticking, thus traps deepen.‡ If the number of kinetic traps is large, as in this model sequence, folding rates will be slowed down at strong HH sticking, leading to a rollover of the chevron plot for folding relaxation.

The chevron plot of a protein that has a sufficiently smooth energy landscape at zero denaturant will have no rollover, as observed experimentally in the kinetics of CI2[53] and CspB.[59,61] Rollover is absent in certain fast phases in the folding of some staphylococcal nuclease mutants.[162] Some rollover may be due to proline isomerization (see Refs. 67 and 68 for the α subunit of tryptophan synthase). In other cases, rollover may result from the diffusion limit on the folding rate, as suggested by the lysozyme mutants studied by B.-L. Chen et al.[71] Chen et al.[71] noted that if the measured rates near the denaturation midpoint is extrapolated linearly, the folding rate at zero denaturant would be of order $10^{-5}s^{-1}$, which is equivalent to a folding time $\approx$ 10 μs. (See also Appendix B for a discussion of "diffusion limits" in lattice models.)

TRANSITION STATES IN THE ENERGY LANDSCAPE PERSPECTIVE: THEY CAN BE LARGE ENSEMBLES

What Do Transition States Explain?

What are the transition states in protein folding? First, what are transition states in general? How they are determined experimentally? At its most fundamental level, the transition state concept is designed to explain why a given process does not go at its fastest possible speed, defined by some reference rate. For example, transition state theory accounts for why chemical bonds form more slowly at finite temperatures than in a single vibration time (k_BT/h in the Eyring theory, h is Planck's constant[174–176]). Transition state theory also addresses why reactions in solution happen slower than the diffusion limit.

For fast-folding, two-state proteins, what a transition state purports to "explain" is that folding is slower under weakly native conditions than under strongly native conditions. This raises two fundamental questions. 1) What is the upper limit for the maximum possible speed a protein could fold? It is

‡Hence, the shapes of model landscapes with ϵ-independent p_{nn} do not scale uniformly with ϵ, as they do in HP and HP+ models. The shape of constant-p_{nn} energy landscapes resembles that of the HP+ model when $|\epsilon/k_BT|$ is small, but in the limit of $\epsilon/(k_BT) \to -\infty$, the energy landscape is identical to that of the corresponding HP sequence.

with respect to this speed that the free energy of activation of a transition state barrier can be computed. 2) What physical process prevents a protein under weakly native conditions from folding as fast as a protein under strongly native conditions?

One possibility, as with small-molecule chemical reactions, is that there is some particular "bottleneck" conformation of the folding protein molecule. A bottleneck conformation would have high internal free energy, corresponding to a saddle point along some reaction pathway on the energy landscape, and be rarely populated. However, according to this scenario, native conditions are expected to cause the protein to achieve this conformation less often than weaker conditions would, in the same way that a high temperature is required to activate a chemical reaction between two reactant atoms. But, in most cases, stronger native conditions in fact speed folding (see below). Nevertheless, we believe that this explanation is likely to be correct for the kinetic phases in which the chains need to escape kinetic traps to reach the native state. For these processes, more *de*stabilizing conditions, such as addition of denaturant, are needed to speed up the kinetics,[163,164] like high temperature in small molecule reactions.

There is also an alternative explanation: the transition state bottleneck sometimes represents a barrier due to the conformational entropy, not a specific high-energy conformation. By definition, two-state kinetics means that there are no intermediates that transiently accumulate during folding. In landscape terms, this means there are no basins where molecules collect, and there are no significant uphill slopes (increase in internal free energy) along the microscopic folding paths on the energy landscape. Yet the transition-state phenomenology appears to be applicable (see, for example, Refs. 63, 65–69, 71, and 177–181). Why does it work?

The landscape picture gives a simple explanation. The folding funnel is shallow under weakly native conditions and is deeper under strongly native conditions. This is because hydrophobicity and other driving forces that determine the internal free energy coordinate of the landscape are weakened by denaturant and strengthened by removal of denaturant. The denatured state can be likened to an ensemble of skiers distributed over the funnel.[8] When folding conditions are "turned on" (denaturant removed), the hydrophobic force is strengthened, stretching the energy landscape so that the native state moves down in internal free energy relative to the denatured states. It is as if turning on folding conditions causes the skiers to all proceed downhill. The reason folding speed depends on denaturant is simply because the degree of landscape stretching is determined by denaturant. Skiers reach the bottom faster on steeper slopes than on shallower slopes. And slopes are steeper under more native-stabilizing conditions.

Hence, the rate limit, or bottleneck, in this process need not be localized on the energy landscape, say to a particular mountain pass. Just as every conformation (except native) is a part of the denatured ensemble, every conformation (except native) contributes to the set that is searched *en route* to the native state. Fast folding resembles the diffusion limit of a diffusing particle: there is no one particular position of a diffusing particle that can be held solely responsible for why the particle cannot move faster. For proteins, this rate limit is a search process over a large conformational ensemble. This is consistent with the observation that the folding rate of the fast-folding protein CspB is sensitive to solvent viscosity.[112] Every element of that ensemble that is partly responsible for the slowness could therefore be called a "transition-state" conformation.

How Free-Energy Barriers Result From Champagne Glass Landscapes

Is such an explanation consistent with experimental data? There are two types of experiments that bear on transition states. One type of experiment monitors folding speed vs. denaturant (or temperature). These are the experiments described above that give chevron plots, which can be interpreted in terms of free-energy barriers. We can use the lattice models to perform the same experiment. We observe relaxation times in the lattice model as a function of driving force, $\epsilon/(k_B T)$. These data are summarized by Eqs. (7) and (9), which can then be represented as standard one-dimensional energy profiles and reaction diagrams (Fig. 24). The lattice model data show a free-energy barrier, just as experiments do, and it can be interpreted by using the same kind of rate law as experiments:

$$k = F \exp\left(-\frac{\Delta G^{\ddagger}}{k_B T}\right) \tag{10}$$

where $\Delta G^{\ddagger}$ is an "activation free energy" and F is a front factor, representing the maximum possible folding speed. Thus, the lattice models are perfectly consistent with *free-energy* barriers in two-state folding kinetics, but these are not barriers in *internal free energy*, the quantity described on energy landscapes.

The second type of experiment that bears on transition states reckons the position of the transition state along the reaction coordinate[54,57,61–63,65,70] (see also Ref. 182). Is the transition state more like the folded state, far forward along the folding reaction coordinate, or like the denatured state, not so far along the reaction coordinate? These experiments are interpreted in terms of a putative "position" of the transition state along the reaction coordi-

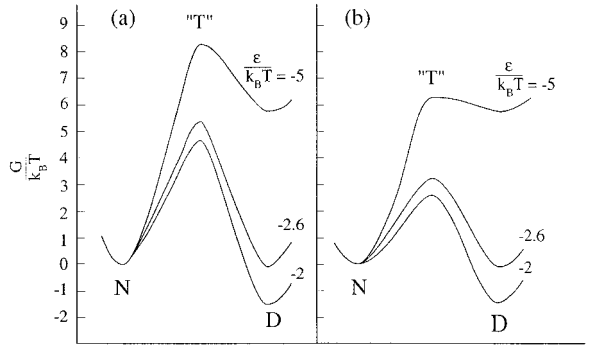

Fig. 24. Computed free energies show typical transition state barriers, in the HP+ model with $\epsilon/(k_BT) = -2$, -2.6, and -5. This is performed by measuring relaxation rates and equilibria vs. folding conditions, ϵ. $\epsilon = -2.6k_BT$ is the equilibrium transition midpoint. N, D, and "T" correspond to the native, unfolded, and a "transition state." The native free energy defines the zero for each profile. The observed activation free energy is affected by the choice of the front factor F. $F = 148.4$ in (**a**), whereas $F = 20.1$ in (**b**).

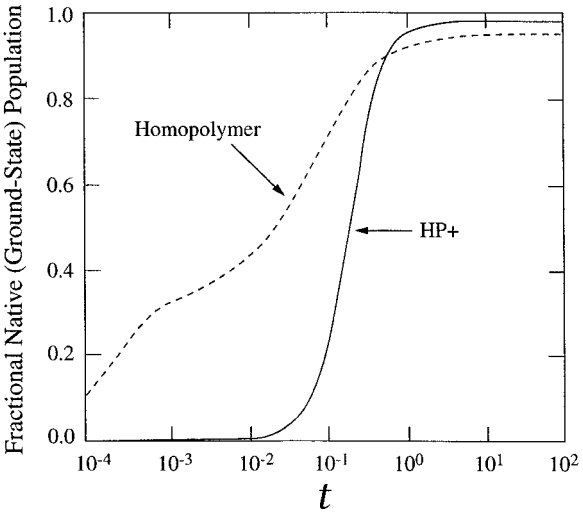

Fig. 25. Homopolymers collapse faster than HP+ chains, computed by using $\epsilon/(k_BT) = -4.17$ with Kawasaki dynamics and the transition matrix method.[37] There are 367 ground-state homopolymer conformations. Populations of the lowest energy states vs. time are shown.

nate. The landscape picture would seem completely incapable of accounting for such positions. If a transition state is an ensemble, and every molecule is following its own trajectory, what is the meaning of a specific point on the reaction coordinate?

Again, our strategy is to use the HP+ model in a completely operational way, collecting data from the model as an experimentalist would in an experiment. When we do this, we find that our model simulations give a result consistent with the experiments. As in the experiments, we take the ratio of the derivatives of folding and unfolding rates[57,63,70] and apply them to the linear relations in Eq. (7) to find that our "transition state" is $-(\partial \ln k_f/\partial\epsilon)/[\partial(-\ln k_u + \ln k_f)/\partial\epsilon] = 1.15/(1.18 + 1.15) = 49\%$ of the way from the native state to the "unfolded state."

Thus, although folding of the HP+ sequence does not have a specific bottleneck structure, nevertheless, we can calculate a "position" of the "transition state" along some undefined reaction coordinate between the unfolded state and the native state. Despite its simplifications, the lattice models produce the same general results as the experiments, but with a picture of the folding transition state that differs from the conventional one. Thus, the success of transition-state phenomenology (Fig. 24) does not necessarily imply a structurally specific bottleneck transition state. Recent simple off-lattice simulations by Guo and Thirumalai[47] also show that the "folding nucleus" of their model protein chains is not unique.

What is the upper "speed limit" for protein folding, F in Eq. (11)? The front factor for gas-phase chemical reactions $(k_BT/h) = 6.2 \times 10^{12}\text{s}^{-1}$ at 25°C (Ref. 183) is not appropriate for protein folding[8,76] (see also the footnote on page 772 of Ref. 178), because folding is not rate limited by the formation of a covalent bond in a vacuum. Recent experiments suggest that the upper limit of contact formation rate is around 10^4s^{-1} to 10^5s^{-1} for small proteins.[184]

What is a good model for the fastest possible folding speed? There are some interesting and subtle issues involved. The HP+ chain folds very quickly and is well directed by a funnel-shaped energy landscape. Folding approximately coincides with collapse. But Figure 25 shows that a homopolymer chain of the same length could collapse[37] even faster to its ground state consisting of maximally compact conformations.§

Why is homopolymer collapse so fast? Because nearly every attempt at making an intrachain contact results in a contact formation and a downhill step on the energy landscape. Also, homopolymers have many native (ground-state) conformations, and any one will do, so this too causes homopolymer collapse to be fast. But when heteropolymers collapse, not all contacts will result in suitable downhill steps. In fact, to facilitate folding to the unique native state, repulsive interactions, such as those in the HP+ model, are needed to eliminate traps. Hence, even folding of a relatively fast-folding heteropolymer sequence is slowed by the need to circumvent hills (see Fig. 14), because it needs to avoid "premature collapse"[3] into traps. Sosnick et al.[58] made a similar point that, during the initial stage of folding, "the great majority of intrachain encounters

§Figure 25 is obtained by using Kawasaki dynamics. The trend is consistent with the calculation using Metropolis dynamics, which indicates that HP+ folding to 80% native population takes more than 6 times longer than the time needed for homopolymer collapse to achieve the same population of maximally compact conformations.

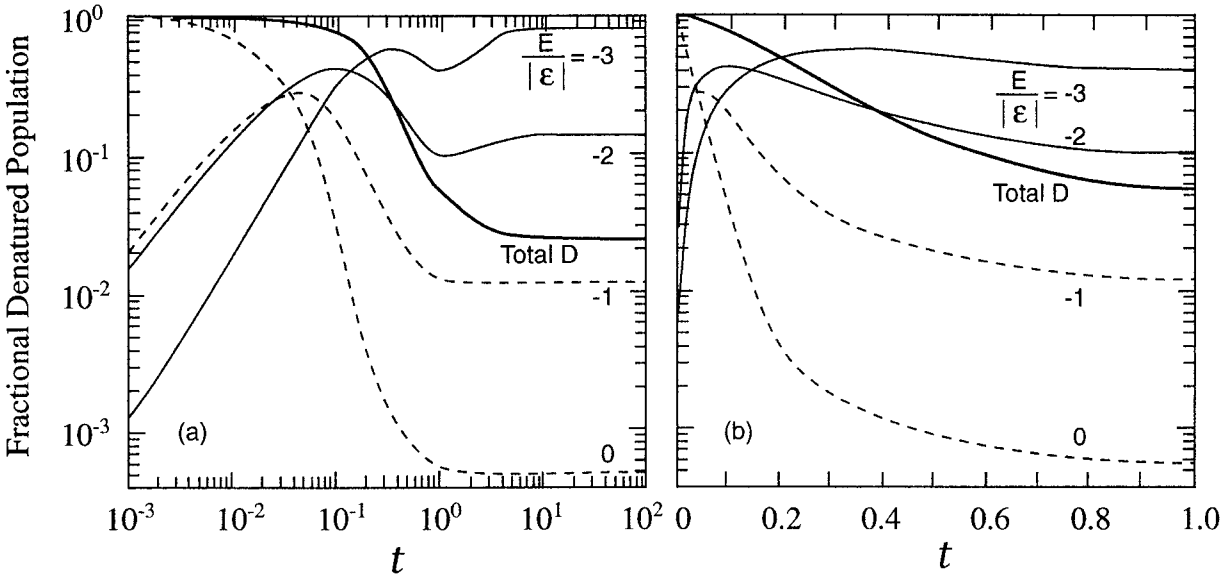

Fig. 26. All denatured subpopulations change with time. There is no preequlibrium during folding of the HP+ sequence, at $\epsilon = -4.17k_BT$, with the same conditions as in Figure 8. Thick solid curves labeled by "Total D" give the population of all denatured (nonnative) conformations as a fraction of the total (i.e., denatured **and** native) chain population. Thin solid and dashed curves give fractions of *only* the denatured conformations in throughway conformational states. Conformational states with direct kinetic connections to the native state are represented by thin solid curves ($E/|\epsilon| = -2, -3$), whereas conformational states without direct kinetic connections to the native state are represented by dashed curves ($E/|\epsilon| = 0, -1$). **a:** Full time range; **b:** Time window when 90% of the folding occurs. The transient decrease in $E/|\epsilon| = -2$ and -3 populations at $t \approx 10^0$ in (a) is due to slowing by kinetic traps.

must fail to achieve productive association and tend to separate."

Another problem of applying transition state theory to protein folding is the idea of "preequilibrium." The conventional transition-state description assumes that, after folding conditions are turned on, the reactant (i.e., denatured conformations) accumulate in a preequilibrium, i.e., in a free-energy basin, from which they can attempt to climb over the free-energy barrier. But our simulations for the HP+ fast-folding funnelscape show no such preequilibrium. Figure 26 shows that populations are continuously changing as folding proceeds. Figure 26a shows that the denatured ensemble does not reach its equilibrium distribution until folding is fully completed. Figure 27 shows that similar features are present in the unfolding kinetics.

It is remarkable that a process as complex as protein folding gives kinetics sometimes as simple as a single exponential. Why does folding sometimes involve single exponentials? Our simulation results shown in Figure 26 would support the following explanation. If the rate of depletion of a given class of nonnative conformations is dependent on the concentration of those conformations and if that concentration is relatively independent of time, an approximate single exponential will result. In our reduced transition matrix formulation, the native state of the HP+ sequence is directly connected (kinetically adjacent) to only two conformational states, the $E/|\epsilon| = -2, -3$ throughway states. Figure 26b shows that, during a substantial fraction of the folding process ($0.2 < t < 0.8$, for example), the *total* populations of these two conformational states *as fractions of the total denatured population* are only weakly time dependent. Thus, we observe approximately single-exponential behavior. It follows that the overall folding rate depends on how quickly the populations of low-energy denatured states ($E/|\epsilon| = -2, -3$) are replenished by downhill transitions from above.

SUMMARY

We have used the HP and HP+ lattice models to explore the energy landscapes of protein folding to interpret experimental data. In the past, one challenge in applying lattice models to experiments has been the difficulty in performing proper ensemble averages over individual folding trajectories. Here because of the model simplicity, we can perform such averages. The HP model tends to give bumpy landscapes with traps, and multiple-exponential dynamics for folding and unfolding. The HP+ model often gives smooth funnel landscapes and approximately single-exponential relaxations. As in experiments, different sequences and conditions can have faster relaxation in the folding than in the unfolding direction, or the reverse, or nearly equal rates. We find that the mechanism of unfolding is not always a direct reversal of the mechanism of folding. Some sequences give chevron plot behavior. We find that

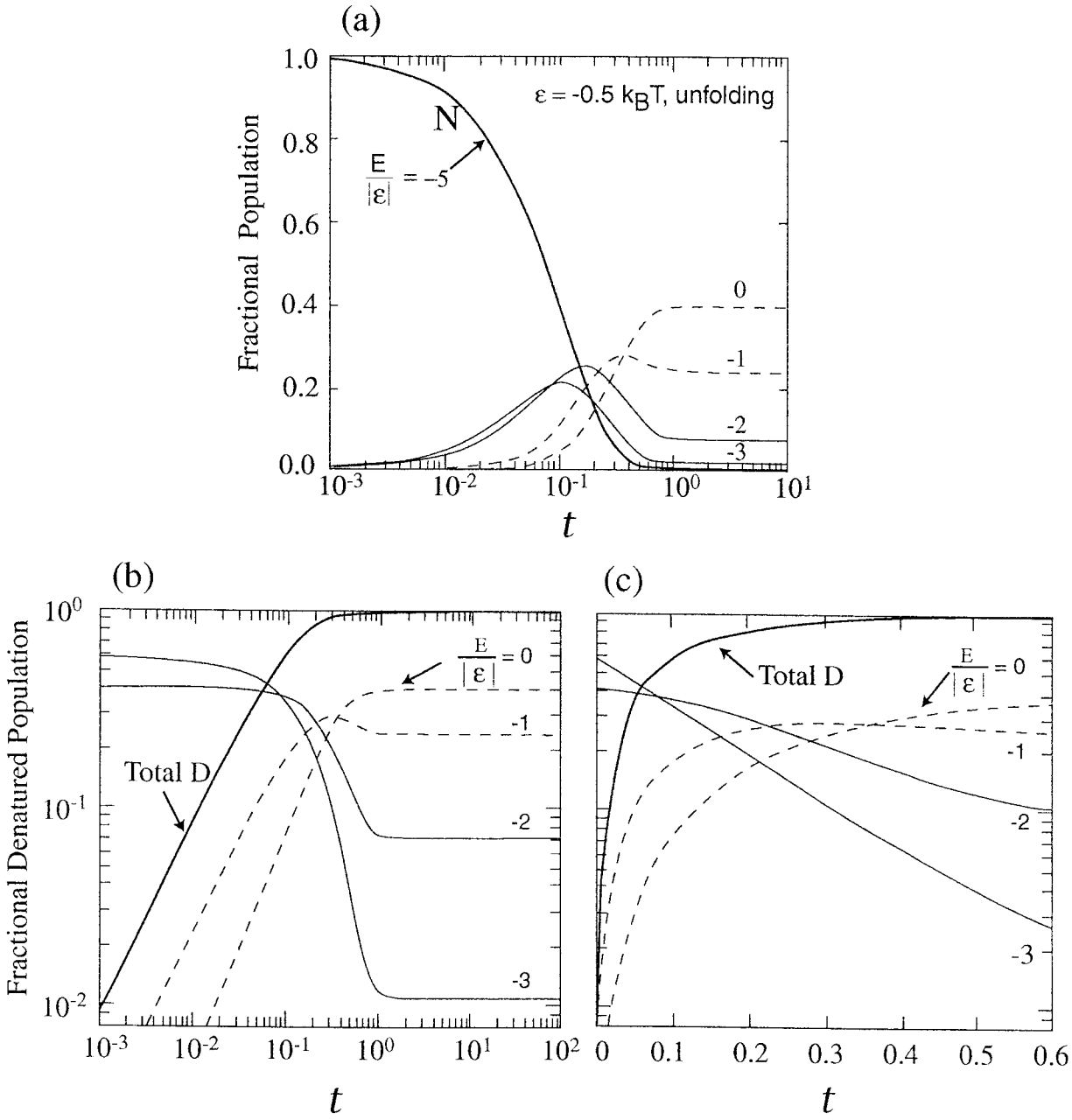

Fig. 27. Time-dependent flows of the subpopulations, for the same HP+ sequence as in Figure 26, now at $\epsilon = -0.5 k_B T$, for the unfolding kinetics. At $t = 0$, all chains begin in the native state N. Numbers indicate the energies of the throughway conformations. Populations of trapped conformations are insignificant during unfolding; $E/|\epsilon| > 0$ populations are not shown. **b, c:** The unfolding counterparts of the folding plots in Figure 26a and b.

the "transition states" in two-state folding are broad ensembles, in which the bottleneck is best described as a slow conformational search process, rather than a specific conformational contortion of the molecule. When we include a reasonable model for the temperature dependence of the hydrophobic interaction, we can reproduce the non-Arrhenius temperature dependence of the experiments of Chen et al.[70] and others. We believe that protein-folding kinetics is often better described in terms of ensembles and energy landscapes than by specific folding pathways of conformational transformations.

NOTE ADDED IN PROOF

Non-Arrhenius behavior has also been observed for the folding rates in a model studied by Karplus et al.[195–197] Since the contact energies in that model are temperature-independent, the non-Arrhenius be-

havior is caused by kinetic traps, similar to that observed in the HP and other models (see, for example, Fig. 10 of Ref. 33, Fig. 20 of Ref. 38, Fig. 4 of Ref. 39, and the ln k_f curve in Fig. 31A in Appendix B of this work). For reasons given in the text, however, we believe that the main cause for non-Arrhenius behaviors in fast-folding, two-state proteins is the temperature dependence of the hydrophobic interaction, rather than the kinetic traps on the energy landscape. Moreover, if model folding rates (k_f) are taken to peak at approximately 25°C, as in experiments, then typical trap mechanisms[38,39,48,82,106,195–197] require temperature ranges in the vicinity of 200K to exhibit significant curvature in the ln k_f vs. $1/T$ plot. The hydrophobic mechanism, on the other hand, leads to significant curvature when temperature is varied over only tens of degrees, in better agreement with experiments.[65,70]

ACKNOWLEDGMENTS

We thank David Agard, David Baker, Robert L. Baldwin, Joe Bryngelson, Chris Dobson, Bill Eaton, Walter Englander, Yuji Goto, Lydia Gregoret, Jim Hofrichter, Sheila Jaswal, Peter Kim, Kunihiro Kuwajima, Bob Matthews, Terry Oas, José Onuchic, Jack Schonbrun, Nick Socci, Julie Sohl, Tobin Sosnick, Dave Thirumalai, Tian Y. Tsong, Zhulun Wang, Peter Wolynes, Kaizhi Yue, and Jill Zitzewitz for helpful discussions. We thank David Agard, David Baker, Chris Dobson, Terry Oas, Franz Schmid, and Dave Thirumalai for making their manuscripts available to us prior to publication. We are very grateful to Danny Heap and Jolanda Schreurs for their creative and technical help with the figures. We thank the National Institutes of Health for financial support.

REFERENCES

1. Baldwin, R.L. Matching speed and stability. Nature 369: 183–184, 1994.
2. Baldwin, R.L. The nature of protein folding pathways: The classical versus the new view. J. Biomol. NMR 5:103–109, 1995.
3. Bryngelson, J.D., Onuchic, J.N., Socci, N.D., Wolynes, P.G. Funnels, pathways and the energy landscape of protein folding: A synthesis. Proteins Struct. Funct. Genet. 21:167–195, 1995.
4. 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.
5. Karplus, M., Šali, A. Theoretical studies of protein folding and unfolding. Curr. Opin. Struct. Biol. 5:58–73, 1995.
6. Wolynes, P.G., Onuchic, J.N., Thirumalai, D. Navigating the folding routes. Science 267:1619–1620, 1995.
7. Thirumalai, D., Woodson, S.A. Kinetics of folding of proteins and RNA. Acc. Chem. Res. 29:433–439, 1996.
8. Dill, K.A., Chan, H.S. From Levinthal to pathways to funnels. Nat. Struct. Biol. 4:10–19, 1997.
9. Anfinsen, C.B., Haber, E., Sela, M., White, F.H., Jr. The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. Proc. Natl. Acad. Sci. USA 47:1309–1314, 1961.
10. Anfinsen, C.B. Principles that govern the folding of protein chains. Science 181:223–230, 1973.
11. Levinthal, C. Are there pathways for protein folding? J. Chim. Phys. 65:44–45, 1968.
12. Levinthal, C. How to fold graciously. In: "Mössbauer Spectroscopy in Biological Systems; Proceedings of a meeting held at Allerton House, Monticello, Illinois." Debrunner, P., Tsibris, J.C.M., Münck, E. (eds.). Urbana, IL: University of Illinois Press, 1969:22–24.
13. Dill, K.A. Theory for the folding and stability of globular proteins. Biochemistry 24:1501–1509, 1985.
14. Zwanzig, R., Szabo, A., Bagchi, B. Levinthal's paradox. Proc. Natl. Acad. Sci. USA 89:20–22, 1992.
15. Dill, K.A. The stabilities of globular proteins. In: "Protein Engineering." Oxender, D.L., Fox, C.F. (eds.). New York: Alan R. Liss, Inc., 1987:187–192.
16. Leopold, P.E., Montal, M., Onuchic, J.N. Protein folding funnels: A kinetic approach to the sequence-structure relationship. Proc. Natl. Acad. Sci. USA 89:8721–8725, 1992.
17. Ikai, A., Tanford, C. Kinetic evidence for incorrectly folded intermediate states in the refolding of denatured proteins. Nature 230:100–102, 1971.
18. Ikai, A., Tanford, C. Kinetics of unfolding and refolding of proteins. I. Mathematical analysis. J. Mol. Biol. 73:145–163, 1973.
19. Ikai, A., Fish, W.W., Tanford, C. Kinetics of unfolding and refolding of proteins. II. Results for cytochrome *c*. J. Mol. Biol. 73:165–184, 1973.
20. Tanford, C., Aune, K.C., Ikai, A. Kinetics of unfolding and refolding of proteins. III. Results for lysozyme. J. Mol. Biol. 73:185–197, 1973.
21. Tsong, T.Y., Baldwin, R.L., Elson, E.L. The sequential unfolding of ribonuclease A: Detection of a fast initial phase in the kinetics of unfolding. Proc. Natl. Acad. Sci. USA 68:2712–2715, 1971.
22. Tsong, T.Y., Baldwin, R.L., Elson, E.L. Properties of the refolding and unfolding reactions of ribonuclease A. Proc. Natl. Acad. Sci. USA 69:1809–1812, 1972.
23. Tsong, T.Y., Baldwin, R.L., McPhie, P., Elson, E.L. A sequential model of nucleation-dependent protein folding: Kinetic studies of ribonuclease A. J. Mol. Biol. 63:453–475, 1972.
24. Tsong, T.Y. Detection of three kinetic phases in the thermal unfolding of ferricytochrome *c*. Biochemistry 12:2209–2214, 1973.
25. Bryngelson, J.D., Wolynes, P.G. Spin glasses and the statistical mechanics of protein folding. Proc. Natl. Acad. Sci. USA 84:7524–7528, 1987.
26. 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.
27. Chan, H.S., Dill, K.A. Intrachain loops in polymers: Effects of excluded volume. J. Chem. Phys. 90:492–509, 1989. Erratum. J. Chem. Phys. 96:3361, 1992.
28. Chan, H.S., Dill, K.A. Compact polymers. Macromolecules 22:4559–4573, 1989.
29. Lau, K.F., Dill, K.A. A lattice statistical mechanics model of the conformational and sequence spaces of proteins. Macromolecules 22:3986–3997, 1989.
30. Flory, P.J. Principles of Polymer Chemistry. Ithaca, NY: Cornell University Press, 1953.
31. Chan, H.S., Dill, K.A. Solvation: How to obtain microscopic energies from partitioning and solvation experiments. Annu. Rev. Biophys. Biomol. Struct. 26:425–459, 1997.
32. Shakhnovich, E.I., Farztdinov, G., Gutin, A.M., Karplus, M. Protein folding bottlenecks: A lattice Monte Carlo simulation. Phys. Rev. Lett. 67:1665–1668, 1991.
33. Miller, R., Danko, C.A., Fasolka, J., Balazs, A.C., Chan, H.S., Dill, K.A. Folding kinetics of proteins and copolymers. J. Chem. Phys. 96:768–780, 1992.
34. Honeycutt, J.D., Thirumalai, D. The nature of folded states of globular proteins. Biopolymers 32:695–709, 1992.
35. Camacho, C.J., Thirumalai, D. Kinetics and thermodynamics of folding in model proteins. Proc. Natl. Acad. Sci. USA 90:6369–6372, 1993.
36. Fukugita, M., Lancaster, D., Mitchard, M.G. Kinematics

and thermodynamics of a folding heteropolymer. Proc. Natl. Acad. Sci. USA 90:6365–6368, 1993.
37. Chan, H.S., Dill, K.A. Energy landscape and the collapse dynamics of homopolymers. J. Chem. Phys. 99:2116–2127, 1993.
38. Chan, H.S., Dill, K.A. Transition states and folding dynamics of proteins and heteropolymers. J. Chem. Phys. 100: 9238–9257, 1994.
39. Socci, N.D., Onuchic, J.N. Folding kinetics of protein-like heteropolymers. J. Chem. Phys. 101:1519–1528, 1994.
40. 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.
41. Thirumalai, D. Theoretical perspectives on in vitro and in vivo protein folding. In: "Statistical Mechanics, Protein Structure, and Protein-Substrate Interactions." Doniach, S. (ed.). New York: Plenum, 1994:115–134.
42. Thirumalai, D., Guo, Z. Nucleation mechanism for protein folding and theoretical predictions for hydrogen-exchange labeling experiments. Biopolymers 35:137–140, 1995.
43. Guo, Z., Thirumalai, D. Kinetics of protein folding: Nucleation mechanism, time scales, and pathways. Biopolymers 36:83–102, 1995.
44. Galzitskaya, O.V., Finkelstein, A.V. Folding of chains with random and edited sequences: Similarities and differences. Protein Eng. 8:883–892, 1995.
45. Zwanzig, R. Simple model of protein folding kinetics. Proc. Natl. Acad. Sci. USA 92:9801–9804, 1995.
46. Veitshans, T., Klimov, D., Thirumalai, D. Protein folding kinetics: Timescales, pathways and energy landscapes in terms of sequence-dependent properties. Folding Design 2:1–22, 1997.
47. Guo, Z., Thirumalai, D. Nucleation-collapse process in protein folding: Evidence for the nonuniqueness of the folding nucleus. Folding & Design. (In press.)
48. Abkevich, V.I., Gutin, A.M., Shakhnovich, E.I. Impact of local and non-local interactions on thermodynamics and kinetics of protein folding. J. Mol. Biol. 252:460–471, 1995.
49. Abkevich, V.I., Gutin, A.M., Shakhnovich, E.I. Specific nucleus as the transition state for protein folding: Evidence from the lattice model. Biochemistry 33:10026–10036, 1994.
50. Shakhnovich, E., Abkevich, V., Ptitsyn, O. Conserved residues and the mechanism of protein folding. Nature 379:96–98, 1996.
51. Thirumalai, D. From minimal models to real proteins: Time scales for protein folding kinetics. J. Physique. I (France) 5:1457–1467, 1995.
52. 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. USA 92:3626–3630, 1995.
53. Jackson, S.E., Fersht, A.R. Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. Biochemistry 30:10428–10435, 1991.
54. Jackson, S.E., Fersht, A.R. Folding of chymotrypsin inhibitor 2. 2. Influence of proline isomerization on the folding kinetics and thermodynamic characterization of the transition state of folding. Biochemistry 30:10436–10443, 1991.
55. Sosnick, T.R., Mayne, L., Hiller, R., Englander, S.W. The barriers in protein folding. Nature Struct. Biol. 1:149–156, 1994.
56. Huang, G.S., Oas, T.G. Structure and stability of monomeric λ repressor: NMR evidence for two-state folding. Biochemistry 34:3884–3892, 1995.
57. Huang, G.S., Oas, T.G. Submillisecond folding of monomeric λ repressor. Proc. Natl. Acad. Sci. USA 92:6878–6882, 1995.
58. Sosnick, T.R., Mayne, L., Englander, S.W. Molecular collapse: The rate-limiting step in two-state cytochrome *c* folding. Proteins Struct. Funct. Genet. 24:413–426, 1996.
59. Schindler, T., Herrler, M., Marahiel, M.A., Schmid, F.X. Extremely rapid protein folding in the absence of intermediates. Nat. Struct. Biol. 2:663–673, 1995.
60. Kiefhaber, T. Kinetic traps in lysozyme folding. Proc. Natl. Acad. Sci. USA 92:9029–9033, 1995.
61. Schindler, T., Schmid, F.X. Thermodynamic properties of an extremely rapid protein folding reaction. Biochemistry 35:16833–16842, 1996.
62. 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. Nat. Struct. Biol. 4:305–310, 1997.
63. Tanford, C. Protein denaturation. Adv. Protein Chem. 23:121–282, 1968.
64. Matthews C.R., Crisanti, M.M., Manz, J.T., Gepner, G.L. Effects of a single amino acid substitution on the folding of the α-subunit of tryptophan synthase. Biochemistry 22: 1445–1452, 1983.
65. Segawa, S.-I., Sugihara, M. Characterization of the transition state of lysozyme unfolding. I. Effect of protein-solvent interactions on the transition state. Biopolymers 23:2473–2488, 1984.
66. Matthews, C.R. Effects of point mutations on the folding of globular proteins. Methods Enzymol. 154:498–511, 1987.
67. Matthews, C.R., Hurle, M.R. Mutant sequences as probes of protein folding mechanisms. BioEssays 6:254–257, 1987.
68. Hurle, M.R., Michelotti, G.A., Crisanti, M.M., Matthews, C.R. Characterization of a slow folding reaction for the α subunit of tryptophan synthase. Proteins Struct. Funct. Genet. 2:54–63, 1987.
69. Chen, B.-L., Schellman, J.A. Low-temperature unfolding of a mutant of phage T4 lysozyme. 1. Equilibrium studies. Biochemistry 28:685–691, 1989.
70. Chen, B.-L., Baase, W.A., Schellman, J.A. Low-temperature unfolding of a mutant of phage T4 lysozyme. 2. Kinetic investigations. Biochemistry 28:691–699, 1989.
71. Chen, B.-L., Baase, W.A., Nicholson, H., Schellman, J.A. Folding kinetics of T4 lysozyme and nine mutants at 12°C. Biochemistry 31:1464–1476, 1992.
72. Matthews, C.R. Pathways of protein folding. Annu. Rev. Biochem. 62:653–683, 1993.
73. Chen, X., Matthews, C.R. Thermodynamic properties of the transition state for the rate-limiting step in the folding of the α subunit of tryptophan synthase. Biochemistry 33:6356–6362, 1994.
74. Matouschek, A., Otzen, D.E., Itzhaki, L.S., Jackson, S.E., Fersht, A.R. Movement of the position of the transition state in protein folding. Biochemistry 34:13656–13662, 1995.
75. Fersht, A.R. Characterizing transition states in protein folding: An essential step in the puzzle. Curr. Opin. Struct. Biol. 5:79–84, 1995.
76. Oliveberg, M., Tan, Y.-J., Fersht, A.R. Negative activation enthalpies in the kinetics of protein folding. Proc. Natl. Acad. Sci. USA 92:8926–8929, 1995.
77. Tan, Y.-J., Oliveberg, M., Fersht, A.R. Titration properties and thermodynamics of the transition state for folding: Comparison of two-state and multi-state folding pathways. J. Mol. Biol. 264:377–389, 1996.
78. Chan, H.S., Dill, K.A. Polymer principles in protein structure and stability. Annu. Rev. Biophys. Biophys. Chem. 20:447–490, 1991.
79. Chan, H.S., Dill, K.A. "Sequence space soup" of proteins and copolymers. J. Chem. Phys. 95:3775–3787, 1991.
80. Chan, H.S., Dill, K.A. The protein folding problem. Physics Today 46:24–32, February 1993.
81. Chan, H.S., Dill, K.A. A simple model of chaperonin-mediated protein folding. Proteins Struct. Funct. Genet. 24:345–351, 1996.
82. Socci, N.D., Onuchic, J.N. Kinetic and thermodynamic analysis of proteinlike heteropolymers: Monte Carlo histogram technique. J. Chem. Phys. 103:4732–4744, 1995.
83. Okazaki, A., Ikura, T., Nikaido, K., Kuwajima, K. The chaperonin GroEL does not recognize apo-α-lactalbumin in the molten globule state. Nat. Struct. Biol. 1:439–446, 1994.
84. Ewbank, J.J., Creighton, T.E., Hayer-Hartl, M.K., Hartl,

F.U. What is the molten globule? Nat. Struct. Biol. 2:10–11, 1995.
85. Taketomi, H., Ueda, Y., Gō, N. Studies on protein folding, unfolding and fluctuations by computer simulation. 1. The effect of specific amino acid sequence represented by specific inter-unit interactions. Int. J. Pept. Protein Res. 7:445–459, 1975.
86. Gō, N., Taketomi, H. Respective roles of short- and long-range interactions in protein folding. Proc. Natl. Acad. Sci. USA 75:559–563, 1978.
87. Abe, H., Gō, N. Noninteracting local-structure model of folding and unfolding transition in globular proteins. II. Application to two-dimensional lattice proteins. Biopolymers 20:1013–1031, 1981.
88. Gupta, P., Hall, C.K. Computer simulation studies of protein refolding pathways and intermediates. Am. Inst. Chem. Eng. J. 41:985–990, 1995.
89. Gō, N., Abe, H. Noninteracting local-structure model of folding and unfolding transition in globular proteins. I. Formulation. Biopolymers 20:991–1011, 1981.
90. O'Toole, E.M., Panagiotopoulos, A.Z. Monte Carlo simulation of folding transitions of simple model proteins using a chain growth algorithm. J. Chem. Phys. 97:8644–8652, 1992.
91. O'Toole, E.M., Panagiotopoulos, A.Z. Effect of sequence and intermolecular interactions on the number and nature of low-energy states for simple model proteins. J. Chem. Phys. 98:3185–3190, 1993.
92. Hao, M.-H., Scheraga, H.A. Monte Carlo simulation of a first-order transition for protein folding. J. Phys. Chem. 98:4940–4948, 1994.
93. Hao, M.-H., Scheraga, H.A. Statistical thermodynamics of protein folding: Sequence dependence. J. Phys. Chem. 98:9882–9893, 1994.
94. Gutin, A.M., Abkevich, V.I., Shakhnovich, E.I. Is burst hydrophobic collapse necessary for protein folding? Biochemistry 34:3066–3076, 1995.
95. Shrivastava, I., Vishveshwara, S., Cieplak, M., Maritan, A., Banavar, J.R. Lattice model for rapidly folding protein-like heteropolymers. Proc. Natl. Acad. Sci. USA 92:9206–9209, 1995.
96. Unger, R., Moult, J. Local interactions dominate folding in a simple protein model. J. Mol. Biol. 259:988–994, 1996.
97. Hao, M.-H., Scheraga, H.A. How optimization of potential functions affects protein folding. Proc. Natl. Acad. Sci. USA 93:4984–4989, 1996.
98. Sun, S., Brem, R., Chan, H.S., Dill, K.A. Designing amino acid sequences to fold with good hydrophobic cores. Protein Eng. 8:1205–1213, 1995.
99. Chan, H.S., Dill, K.A. Comparing folding codes for proteins and polymers. Proteins Struct. Funct. Genet. 24:335–344, 1996.
100. 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.
101. Šali, A., Shakhnovich, E., Karplus, M. How does a protein fold? Nature 369:248–251, 1994.
102. Shakhnovich, E.I. Proteins with selected sequences fold into unique native conformation. Phys. Rev. Lett. 72:3907–3910, 1994. Erratum. Phys. Rev. Lett. 74:2618, 1995.
103. Shakhnovich, E.I., Gutin, A.M. Implications of thermodynamics of protein folding for evolution of primary sequences. Nature 346:773–775, 1990.
104. Gō, N. Theoretical studies of protein folding. Annu. Rev. Biophys. Bioeng. 12:183–210, 1983.
105. Pande, V.S., Grosberg, A. Yu, Tanaka, T. On the theory of folding kinetics for short proteins. Folding Design 2:109–114, 1997.
106. 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.
107. Kuntz, I.D., Crippen, G.M., Kollman, P.A., Kimelman, D. Calculation of protein tertiary structure. J. Mol. Biol. 106:983–994, 1976.
108. Kawasaki, K. Diffusion constants near the critical point for time-dependent Ising models. I. Phys. Rev. 145:224–230, 1966.
109. Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E. Equation of state calculations by fast computing machines. J. Chem. Phys. 21:1087–1092, 1953.
110. Klimov, D.K., Thirumalai, D. A criterion that determines the foldability of proteins. Phys. Rev. Lett. 76:4070–4073, 1996.
111. Klimov, D.K., Thirumalai, D. Factors governing the foldability of proteins. Proteins Struct. Funct. Genet. 26:411–441, 1996.
112. Jacob, M., Schindler, T., Balbach, J., Schmid, F.X. Diffusion control in an elementary protein folding reaction. Proc. Natl. Acad. Sci. USA. 94:5622–5627, 1997.
113. Kim, P.S., Baldwin, R.L. Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. Annu. Rev. Biochem. 51:459–489, 1982.
114. Schmid, F.X. Mechanism of folding ribonuclease A. Slow refolding is a sequential reaction via structural intermediates. Biochemistry 22:4690–4696, 1983.
115. Chan, H.S. Kinetics of protein folding. Nature 373:664–665, 1995.
116. Camacho, C.J., Thirumalai, D. Modeling the role of disulfide bonds in protein folding: Entropic barriers and pathways. Proteins Struct. Funct. Genet. 22:27–40, 1995.
117. Radford, S.E., Dobson, C.M., Evans, P.A., The folding of hen lysozyme involves partially structured intermediates and multiple pathways. Nature 358:302–307, 1992.
118. Dobson, C.M., Evans, P.A., Radford, S.E. Understanding how proteins fold: The lysozyme story so far. Trends Biochem Sci. 19:31–37, 1994.
119. Hill, T.L. Effect of rotation on the diffusion-controlled rate of ligand-protein association. Proc. Natl. Acad. Sci. USA 72:4918–4922, 1975.
120. Hill, T.L. Diffusion frequency factors in some simple examples of transition-state rate theory. Proc. Natl. Acad. Sci. USA 73:679–683, 1976.
121. Zhou, H.-X., Zwanzig, R. A rate process with an entropy barrier. J. Chem. Phys. 94:6147–6152, 1991.
122. Hekman, C.M., Loskutoff, D.J. Endothelial cells produce a latent inhibitor of plasminogen activators that can be activated by denaturants. J. Biol. Chem. 260:11581–11587, 1985.
123. Carrell, R.W., Evans, D.L., Stein, P.E. Mobile reactive centre of serpins and the control of thrombosis. Nature 353:576–578, 1991. Erratum Nature 364:737, 1993.
124. Mottonen, J., Strand, A., Symersky, J., Sweet, R.M., Danley, D.E., Geoghegan, K.F., Gerard, R.D., Goldsmith, E.J. Structural basis of latency in plasminogen activator inhibitor-1. Nature 355:270–273, 1992.
125. Wang, Z., Mottonen, J., Goldsmith, E.J. Kinetically controlled folding of the serpin plasminogen activator inhibitor 1. Biochemistry 35:16443–16448, 1996.
126. Sohl, J.L., Jaswal, S.S., Agard, D.A. The mechanism of kinetic control in α-lytic protease. Submitted.
127. Brandts, J.F., Halvorson, H.R., Brennan, M. Consideration of the possibility that the slow step in protein denaturation reactions is due to cis-trans isomerism of proline residues. Biochemistry 14:4953–4963, 1975.
128. Lin, L.-N., Brandts, J.F. Further evidence suggesting that the slow phase in protein unfolding and refolding is due to proline isomerization: A kinetic study of carp parvalbumins. Biochemistry 17:4102–4110, 1978.
129. Schmid, F.X., Baldwin, R.L. Acid catalysis of the formation of the slow-folding species of RNase A: Evidence that the reaction is proline isomerization. Proc. Natl. Acad. Sci. USA 75:4764–4768, 1978.
130. Kim, P.S., Baldwin, R.L. Intermediates in the folding reactions of small proteins. Annu. Rev. Biochem. 59:631–660, 1990.
131. Nall, B.T. Proline isomerization as a rate-limiting step. In: "Mechanisms of Protein Folding." Pain, R.H. (ed.). New York: Oxford University Press, 1994:80–103.
132. Hagerman, P.L. Kinetic analysis of the reversible folding reactions of small proteins: Application to the folding of

lysozyme and cytochrome c. Biopolymers 16:731–747, 1977.
133. Ridge, J.A., Baldwin, R.L., Labhardt, A.M. Nature of the fast and slow refolding reactions of iron(III) cytochrome c. Biochemistry 20:1622–1630, 1981.
134. Udgaonkar, J.B., Baldwin, R.L. Early folding intermediate of ribonuclease. A. Proc. Natl. Acad. Sci. USA 87:8197–8201, 1990.
135. Herning, T., Yutani, K. Taniyama, Y., Kikuchi, M. Effects of proline mutations on the unfolding and refolding of human lysozyme: The slow refolding kinetic phase does not result from proline cis-trans isomerization. Biochemistry 30:9882–9891, 1991.
136. Kiefhaber, T., Grunert, H.-P., Hahn, U., Schmid, F.X. Folding of RNase T1 is decelerated by a specific tertiary contact in the folding intermediate. Proteins Struct. Funct. Genet. 12:171–179, 1992.
137. Creighton, T.E. The energetic ups and downs of protein folding. Nat. Struct. Biol. 1:135–138, 1994.
138. Mann, C.J., Shao, X., Matthews, C.R. Characterization of the slow folding reactions of *trp* aporepressor from *Escherichia coli* by mutational analysis of prolines and catalysis by a peptidyl-prolyl isomerase. Biochemistry 34:14573–14580, 1995.
139. Kuwajima, K., Hiraoka, Y., Ikeguchi, M., Sugai, S. Comparison of the transient folding intermediates in lysozyme and α-lactalbumin. Biochemistry 24:874–881, 1985.
140. Ikeguchi, M., Kuwajima, K., Mitani, M., Sugai, S. Evidence for identity between the equilibrium unfolding intermediate and a transient folding intermediate: A comparative study of the folding reactions of α-lactalbumin and lysozyme. Biochemistry 25:6965–6972, 1986.
141. Balbach, J., Forge, V., van Nuland, N.A., Winder, S.L., Hore, P.J., Dobson, C.M. Following protein folding in real time using NMR spectroscopy. Nat. Struct. Biol. 2:865–870, 1995.
142. Roder, H. Watching protein folding unfold. Nat. Struct. Biol. 2:817–820, 1995.
143. Kuwajima, K. The molten globule state as a clue for understanding the folding and cooperativity of globular-protein structure. Proteins Struct. Funct. Genet. 6:87–103, 1989.
144. Baldwin, R.L. Pulsed H/D-exchange studies of folding intermediates. Curr. Opin. Struct. Biol. 3:84–91, 1993.
145. Weissman, J.S. All roads lead to Rome? The multiple pathways of protein folding. Curr. Biol. 2:255–260, 1995.
146. Kuwajima, K., Yamaya, H., Miwa, S., Sugai, S., Nagamura, T. Rapid formation of secondary structure framework in protein folding studied by stopped-flow circular dichroism. FEBS Lett. 221:115–118, 1987.
147. Hamada, D., Kuroda, Y., Tanaka, T., Goto, Y. High helical propensity of the peptide fragments derived from β-lactoglobulin, a predominantly β-sheet protein. J. Mol. Biol. 254:737–746, 1995.
148. Shiraki, K., Nishikawa, K., Goto, Y. Trifluoroethanol-induced stabilization of the α-helical structure of β-lactoglobulin: Implication for non-hierarchical protein folding. J. Mol. Biol. 245:180–194, 1995.
149. Hamada, D., Segawa, S.-I., Goto, Y. Non-native α-helical intermediates in the refolding of β-lactoglobulin, a predominantly β-sheet protein. Nat. Struct. Biol. 3:868–873, 1996.
150. Kuroda, Y., Hamada, D., Tanaka, T., Goto, Y. High helicity of peptide fragments corresponding to β-strand regions of β-lactoglobulin observed by 2D-NMR spectroscopy. Folding Design 1:243–251, 1996.
151. Kuwajima, K., Yamaya, H., Sugai, S. The burst-phase intermediate in the refolding of β-lactoglobulin studied by stopped-flow circular dichroism and absorption spectroscopy. J. Mol. Biol. 264:806–822, 1996.
152. Itzhaki, L.S., Evans, P.A., Dobson, C.M., Radford, S.E. Tertiary interactions in the folding pathway of hen lysozyme: Kinetic studies using fluorescent probes. Biochemistry 33:5212–5220, 1994.
153. Chaffotte, A.F., Guillou, Y., Goldberg, M.E. Kinetic resolution of peptide bond and side chain far-uv circular dichroism during the folding of hen egg white lysozyme. Biochemistry 31:9694–9702, 1992.
154. Roder, H., Elöve, G.A., Englander, S.W. Structural characterization of folding intermediates in cytochrome *c* by H-exchange labeling and proton NMR. Nature 335:700–704, 1988.
155. Bycroft, M., Matouschek, A., Kellis, J.T., Serrano, L., Fersht, A.R. Detection and characterization of a folding intermediate in barnase by NMR. Nature 346:488–490, 1990.
156. Alonso, D.O.V., Dill, K.A. Solvent denaturation and stabilization of globular proteins. Biochemistry 30:5974–5985, 1991.
157. Shortle, D., Chan, H.S., Dill, K.A. Modeling the effects of mutations on the denatured states of proteins. Protein Sci. 1:201–215, 1992.
158. 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.
159. Camacho, C.J., Thirumalai, D. Denaturants can accelerate folding rates in a class of globular proteins. Protein Sci. 5:1826–1832, 1996.
160. McCoy, L.F., Rowe, E.S., Wong, K.P. Multiparameter kinetic study on the unfolding and refolding of bovine carbonic anhydrase B. Biochemistry 19:4738–4743, 1980.
161. 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, 1993.
162. Kalnin, N.N., Kuwajima, K. Kinetic folding and unfolding of staphylococcal nuclease and its six mutants studied by stopped-flow circular dichroism. Proteins Struct. Funct. Genet. 23:163–176, 1995.
163. Goto, Y., Hamaguchi, K. Formation of the intrachain disulfide bond in the constant fragment of the immunoglobulin light chain. J. Mol. Biol. 146:321–340, 1981.
164. Weissman, J.S., Kim, P.S. Reexamination of the folding of BPTI: Predominance of native intermediates. Science 253:1386–1393, 1991.
165. Santoro, M.M., Bolen, D.W. A test of the linear extrapolation of unfolding free energy changes over an extended denaturant concentration range. Biochemistry 31:4901–4907, 1992.
166. Yao, M., Bolen, D.W. How valid are denaturant-induced unfolding free energy measurements? Level of conformance to common assumptions over an extended range of ribonuclease A stability. Biochemistry 34:3771–3781, 1995.
167. Dill, K.A., Alonso, D.O.V., Hutchinson, K. Thermal stabilities of globular proteins. Biochemistry 28:5439–5449, 1989.
168. Alexander, P., Orban, J., Bryan, P. Kinetic analysis of folding and unfolding the 56 amino acid IgG-binding domain of streptococcal protein G. Biochemistry 31:7243–7248, 1992.
169. 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. USA. 94:10636–10640, 1997.
170. Baldwin, R.L. On-pathway versus off-pathway folding intermediates. Folding Design 1:R1–R8, 1996.
171. Matouschek, A., Kellis, J.T., Jr., Serrano, L., Bycroft, M., Fersht, A.R. Transient folding intermediates characterized by protein engineering. Nature 346:440–445, 1990.
172. 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. Nat. Struct. Biol. 3:193–205, 1996.
173. Chen, H.M., Markin, V.S., Tsong, T.Y. Kinetic evidence of microscopic states in protein folding. Biochemistry 31: 12369–12375, 1992.
174. Laidler, K.J. "Chemical Kinetics." 2nd edit., New York: McGraw-Hill, 1965:72–112.
175. Eyring, H., Lin, S.H., Lin, S.M. "Basic Chemical Kinetics." New York: Wiley, 1980:235–258.
176. Steinfeld, J.I., Francisco, J.S., Hase, W.L. Chemical Kinet-

TABLE I. Conformational Distance From Native δ vs. Order Parameter *Q*, for the HP Sequence

	δ												Average
Q	0	1	2	3	4	5	6	7	8	9	10	11	δ
0	0	0	2	159	1,703	7,052	10,407	4,817	633	18	0	0	5.8
1/6	0	0	43	548	2,444	4,535	3,245	1,249	553	87	25	7	5.3
1/3	0	2	85	436	821	709	384	174	80	56	17	1	4.8
1/2	0	12	47	57	57	51	31	19	9	7	5	1	4.3
2/3	0	3	5	6	4	4	0	1	1	2	2	0	4.3
5/6	0	0	0	0	0	0	0	0	0	0	0	0	—
1	1	0	0	0	0	0	0	0	0	0	0	0	0

Numbers of conformations are classified by δ and *Q*.

ics and Dynamics. Englewood Cliffs, NJ: Prentice-Hall, 1989:402–428.
177. Eyring, H., Stearn, A.E. The application of the theory of absolute reaction rates to proteins. Chem. Rev. 24:253–270, 1939.
178. Fersht, A.R., Matouschek, A., Serrano, L. The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. J. Mol. Biol. 224:771–782, 1992.
179. Serrano, L., Matouschek, A., Fersht, A.R. The folding of an enzyme. III. Structure of the transition state for unfolding of barnase analysed by a protein engineering procedure. J. Mol. Biol. 224:805–818, 1992.
180. Jackson, S.E., elMasry, N., Fersht, A.R. Structure of the hydrophobic core in the transition state for folding of chymotrypsin inhibitor 2: A critical test of the protein engineering method of analysis. Biochemistry 32:11270–11278, 1993.
181. Fersht, A.R. Optimization of rates of protein folding: The nucleation-condensation mechanism and its implications. Proc. Natl. Acad. Sci. USA 92:10869–10873, 1995.
182. 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.
183. Creighton, T.E. "Proteins: Structures and Molecular Properties." 2nd edit. New York: W.H. Freeman, 1993:180–182.
184. Hagen, S.J., Hofrichter, J., Szabo, A., Eaton, W.A. Diffusion-limited contact formation in unfolded cytochrome *c*: Estimating the maximum rate of protein folding. Proc. Natl. Acad. Sci. USA 93:11615–11617, 1996.
185. Kanehisa, M.I., Tsong, T.Y. Mechanisms of the multiphasic kinetics in the folding and unfolding of globular proteins. J. Mol. Biol. 124:177–194, 1978.
186. Doyle, R., Simons, K., Qian, H., Baker, D. Local interactions and the optimization of protein folding. Proteins 29:282–291, 1997.
187. 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.
188. Hansen, J.P., MacDonald, I.R. "Theory of Simple Liquids." 2nd edit. London: Academic Press Inc., 1986:57–61.
189. 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.
190. Waldburger, C.D., Jonsson, T., Sauer, R.T. Barriers to protein folding: Formation of buried polar interactions is a slow step in acquisition of structure. Proc. Natl. Acad. Sci. USA 93:2629–2634, 1996.
191. Baldwin, R.L. Why is protein folding so fast? Proc. Natl. Acad. Sci. USA 93:2627–2628, 1996.
192. Klimov, D.K., Thirumalai, D. Viscosity dependence of the folding rates of proteins. Phys. Rev. Lett. 79:317–320, 1997.
193. Karplus, M., Šali, A., Shakhnovich, E. Reply to "Kinetics of protein folding." Nature 373:665, 1995.
194. Shakhnovich, E.I. Theoretical studies of protein-folding thermodynamics and kinetics. Curr. Opin. Struct. Biol. 7:29–40, 1997.
195. Karplus, M., Caflisch, A., Šali, A., Skakhnovich, E. Protein Dynamics: From the native to the unfolded state and back again. In: "Modelling of Biomolecular Structures and Mechanisms. Proceedings of the Twenty-Seventh Jerusalem Symposium on Quantum Chemistry and Biochemistry." Pullman, A., Jortner, J., Pullman, B. (eds.). Boston: Kluwer Academic Publishers, 1994:69–84.
196. Karplus, M. The Levinthal paradox: yesterday and today. Folding & Design 2:S69–S75, 1997.
197. Dobson, C.M. Šali, A., Karplus, M. Protein folding: A perspective from theory and experiment. Angewandte Chemie. In Press.

APPENDIX A
REACTION COORDINATES: δ AND Q

The thermodynamic order parameter *Q* is defined as the number of native contacts in a given conformation divided by the total number of native contacts in the native conformation. ($Q = 1$ corresponds to the native state.) Recently, *Q* was used as a "reaction coordinate" to study folding kinetics in lattice models.[6,52,101,106,182]¶ Here we compare *Q* with the conformational distance δ (our kinetic reaction coordinate, defined in the text, Ref. 38). The correlation between *Q* and δ are given by the numbers of conformations in different (*Q*, δ) categories in Tables I and II for the HP and HP+ sequences in Figure 1.

The correlation between *Q* and δ is weak. For $Q = \frac{1}{6}$ possible values for δ cover almost the entire range of possible values for δ (Tables 1 and 2). *Q* does not always reflect kinetic progress toward the native state. For example, along pathway III in Figure 6, the trapped conformation B has $\delta = 7$, and it has to overcome an energetic barrier before it can reach the native state. Yet its *Q* value is ½, which is higher than the throughway conformations at $\delta = 4$, 5, and 6 along the same path, all of which have $Q = 0$.

Q contains no notion of kinetic barriers or move set of the model kinetics. Real chains cannot take large leaps through conformational space in a single time step; this is taken into account in kinetic reaction

¶The parameter *Q* used here is equivalent to Q_0 in Ref. 101 and is equivalent to the unnormalized *Q* in Refs. 52, 106, and 182 divided by the number of native contacts. *Q* is also used as the single reaction coordinate in a recent analytical model (Ref. 105).

TABLE II. Conformational Distance From Native δ vs. Order Parameter *Q*, for the HP+ Sequence

	δ										Average
Q	0	1	2	3	4	5	6	7	8	9	δ
0	0	0	2	159	1,631	6,818	10,600	4,971	594	16	5.8
1/6	0	0	43	532	2,349	4,636	3,896	1,163	114	3	5.2
1/3	0	2	85	411	784	800	515	151	16	1	4.6
1/2	0	12	43	53	53	71	48	14	2	0	4.1
2/3	0	3	5	4	8	6	2	0	0	0	3.5
5/6	0	0	0	0	0	0	0	0	0	0	—
1	1	0	0	0	0	0	0	0	0	0	0

Numbers of conformations are classified by δ and *Q*.

coordinates, such as δ, which recognize that only certain conformational transitions are permitted in a time step, but such limitations are not taken into account in thermodynamic coordinates, such as *Q*. Conversely, a small conformational change can result in a large change in *Q*. For example, in the ($\delta = 7$) ⇌ ($\delta = 6$) transition along pathway III in Figure 6, ($Q = 1/3$) ⇌ ($Q = 0$).

Some model studies use *Q* as a reaction coordinate.[6,52,101,105,106,182] But the shapes of energy profiles vs. *Q* are not necessarily related to the kinetics of a model. This has been pointed out by Socci et al.[106] Figure 28 shows that such a reaction coordinate fails to correctly capture the kinetics. For strong HH sticking conditions, a "barrier" develops at $Q = 2/3$ for the HP+ sequence (Fig. 28B) but not for the HP sequence (Fig. 28A), which may be taken to imply that the HP+ chain should fold more slowly. In fact, the actual situation is the opposite. Furthermore, in neither energy function is there any conformation with exactly five native contacts. Therefore, there is an infinite "barrier" at $Q = 5/6$. But, in fact, no such barrier to folding really exists: a single move can readily cause more than one contact to form at a time. Hence, neither *Q* nor other thermodynamic "order parameters" are proper reaction coordinates in general. Similar problems may appear in other reaction coordinates[45,185,186] that do not account for steric and other constraints on chain movements.

An additional problem is the use of $Q = 1/2$ to define the equilibrium midpoints of transitions.[40,48,49,94,102,187] But more correctly, the equilibrium transition midpoint is where one half of the chain population adopts the unique native conformation.[38,39,81,82,157] Figure 29 shows the error in using this prescription. Therefore, contrary to an earlier assertion,[102] the equilibrium *Q* value is not equal to the time the chain spends in the native state.

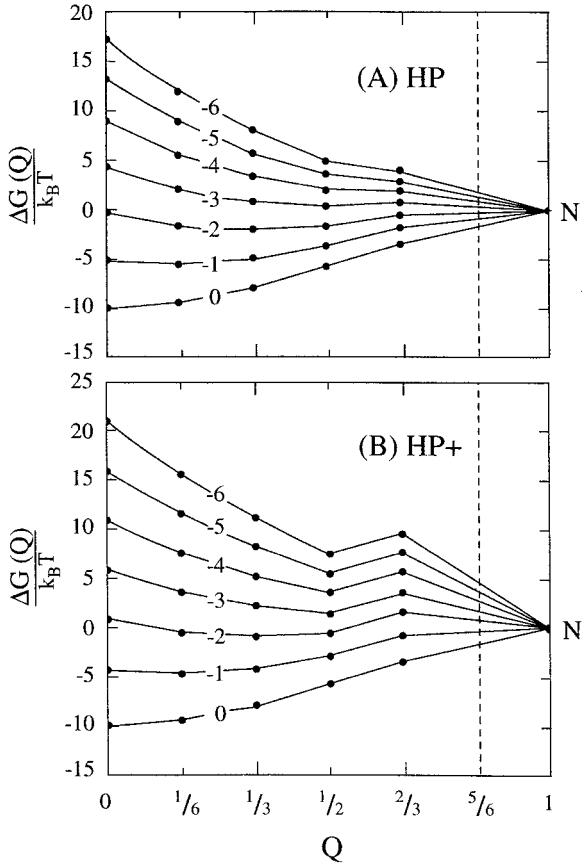

Fig. 28. Free energy of the chains vs. *Q*, fractional number of native contacts. $\Delta G(Q) \equiv G(Q) - E_N$ is the free energy relative to the native (N) energy for (**A**) the HP and (**B**) the HP+ sequences in Figure 1. Labels are $\epsilon/(k_B T)$. For both sequences, the vertical dashed lines indicate that there is no conformation with $Q = 5/6$.

APPENDIX B
METROPOLIS VS. KAWASAKI DYNAMICS

Metropolis Monte Carlo methods[109] were designed for studying equilibrium properties of matter. They have also been applied to studying kinetics. Their application to protein-folding kinetics began with the pioneering work of Taketomi et al.[85] Here, we discuss why Kawasaki sampling[108] in Monte Carlo algorithms may often be more appropriate than Metropolis sampling for folding kinetics.

In the transition matrix formulation of model protein kinetics,[37,38] the matrix element $(\mathbf{T})_{ab}$ is the probability of a transition from conformation *a* (energy E_a) to conformation *b* (energy E_b). The Metropo-

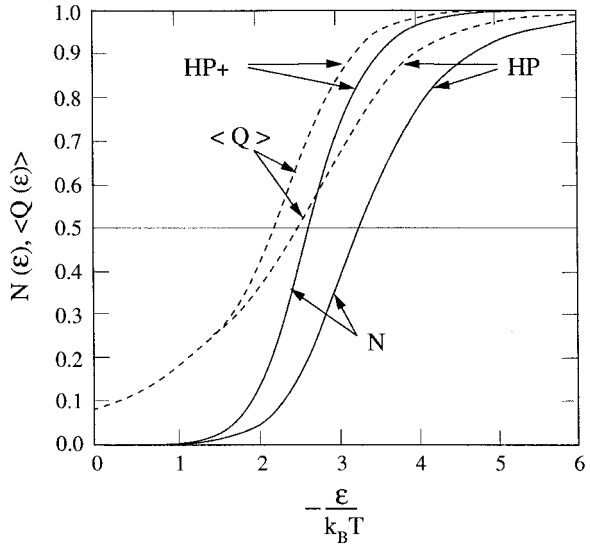

Fig. 29. $\langle Q \rangle = \frac{1}{2}$ does not define the transition midpoint. True equilibrium native populations of the model (N, solid curves) and the Boltzmann-averaged Q-parameter ($\langle Q \rangle$, dashed curves) for the HP and HP+ sequences in Figure 1 vs. $\epsilon/(k_B T)$.

lis criterion for the probability of accepting a transition[109] is given by

$$(\mathbf{T})_{ab} = \begin{cases} \min\{1, \exp[-\Delta E_{ab}/(k_B T)]\} A(a,b)/A & \text{for } a \neq b \\ 1 - \Sigma_{b \neq a} (\mathbf{T})_{ab} & \text{for } a = b \end{cases} \tag{11}$$

where $\Delta E_{ab} \equiv E_b - E_a$ is the energy difference between the two conformations. $A(a, b) = 1$ if the two conformations are adjacent; otherwise $A(a, b) = 0$. The denominator A is a normalization constant (denoted by $A_{\max}$ in Refs. 37, 38, and 81).

The gradient of an energy landscape is an important determinant of the speed of moving uphill or downhill. But the Metropolis criterion only recognizes the uphill gradient (i.e., sampling is rarer as $\Delta E_{ab} > 0$ increases). All downhill steps ($\Delta E_{ab} \leq 0$) are accepted with the same probability. This corresponds to the physical assumption that the spatial range of favorable contact interaction is literally zero, so monomers along the chain would not feel any attraction to form a favorable contact.

A more symmetric sampling criterion is due to Kawasaki.[108] The transition probability in a generalized Kawasaki model is given by

$$(\mathbf{T})_{ab} = f(|\Delta E_{ab}|) \exp\left[-\frac{\Delta E_{ab}}{2k_B T}\right] \frac{A(a, b)}{A} \tag{12}$$

for $a \neq b$, and $(\mathbf{T})_{aa} = 1 - \Sigma_{b \neq a} (\mathbf{T})_{ab}$ for $a = b$. Here, the denominator A is a sufficiently large normalization constant so that all transition probabilities are nonnegative. For any even function $f(\Delta E_{ab}) = f(-\Delta E_{ab})$, the Kawasaki dynamics Eq. (12) leads to the Boltzmann distribution at equilibrium because it satisfies microscopic reversibility.[188] We study a modified Kawasaki dynamics using

$$f(|\Delta E_{ab}|) = \left[\cosh\left(\frac{\Delta E_{ab}}{2k_B T}\right)\right]^{-\gamma}. \tag{13}$$

The original algorithm of Kawasaki[108] corresponds to $\gamma = 1$.

In Kawasaki dynamics, in contrast to Metropolis sampling, the larger the downhill gradient $\Delta E_{ab} < 0$, the greater the probability of sampling. The degree of this downhill directedness is modeled by the parameter γ in Eq. (13), where smaller γ (including $\gamma < 0$) imply faster rates toward low-energy states. We use $\gamma = 0$ to obtain the results in the text.

Time evolution in the transition matrix formulation is obtained by raising the transition matrix $\mathbf{T}$ to successively higher powers, $\mathbf{T}^\tau$ (Refs. 37 and 38). The model time t is defined by $t \equiv \tau/A$ (Ref. 81). For sufficiently large A , rates expressed in terms of t are independent of A. In the $A \to \infty$ limit, the transition matrix formulation reduces to the standard differential master equation formulation. The exact unfolding rate in Figure 21 is obtained in this limit. (The rest of the data points in Fig. 21 are obtained with $A = 10^6$.) For computational efficiency, we do not use exceedingly large values of A. $A = 29$ is used for the MS2 Metropolis dynamics calculation in this Appendix. $A = 10^6$ is used in Figure 23, and $A = 10^5$ is used in Figure 27. In all other Kawasaki folding or unfolding calculations in this work, we use a value of A that is 10% larger than the maximum probability of transition out of any conformational state in the reduced matrix formulation for the given ϵ. The same A as the HP+ kinetics in Figure 8 is used for the homopolymer collapse kinetics in Figure 25.

We have compared Kawasaki and Metropolis dynamics in a few situations. We found that the general trends are the same and that there were no cases where the choice of sampling method changes multiple-exponential to single-exponential dynamics, or vice versa. Figure 30 shows the Metropolis relaxation curves in the folding direction for both sequences at different HH sticking energies. Rates estimated from these curves are plotted in Figure 31 as functions of ϵ.

The Metropolis Figure 31A is different from the Kawasaki HP chevron plot in Figure 17 because the major Metropolis kinetic phase involves escaping

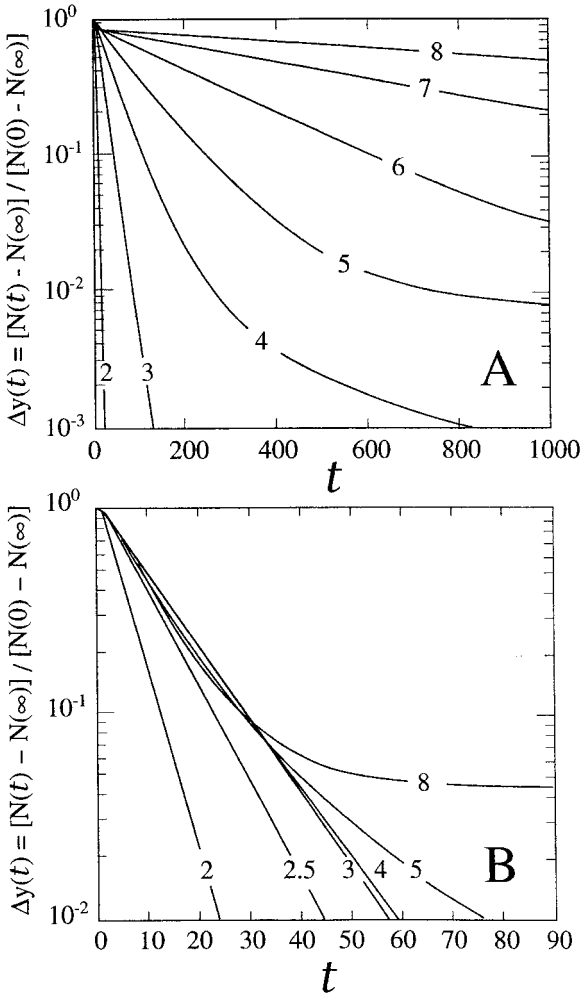

Fig. 30. Metropolis folding relaxation curves for the sequence in Figure 1 using (**A**) the HP model and (**B**) the HP+ model. The labels are $-\epsilon/(k_B T)$.

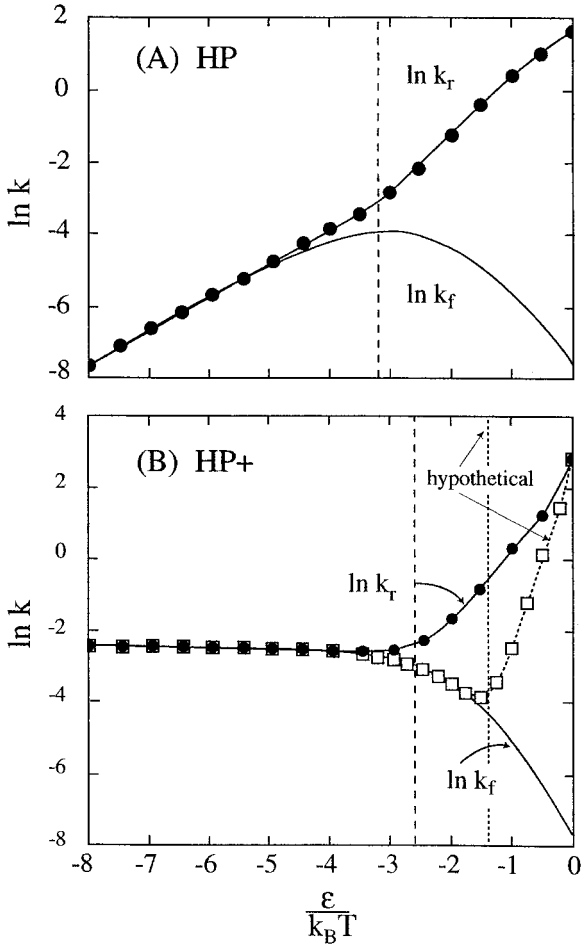

Fig. 31. Relaxation rates ln k_f vs. ϵ in Metropolis dynamics, for sequence in Figure 1, for (**A**) the HP and (**B**) the HP+ model (filled circles). Folding rates (ln k_f , solid curves) are computed by reduced transition matrices for first-passage times.[38] The values for k_f are estimated by the time needed for the fraction of chains that have passed through the native state changes from approximately 0.2 to 0.7. Squares in (B) show the folding relaxation rates (ln k_f) using Metropolis dynamics for a hypothetical energy landscape, identical to that of the HP+ model except that its native energy is artificially lowered from the HP+ sequence's $E_N = 5\epsilon$ (Fig. 2) to $E_N = 8\epsilon$. The vertical dashed lines indicate the equilibrium folding-denaturation transition midpoints of the energy landscapes considered.

from energy traps (≈80% at strong HH sticking strength). Hence, the folding relaxation rate of the major kinetic phase decreases monotonically as HH sticking becomes stronger, similar to the minor slow kinetic phase in Kawasaki dynamics (Fig. 20). Figure 31A also shows the folding rate k_f computed by Metropolis first-passage times. The overall shape of the ln k_f curve has a maximum rate at approximately $\epsilon = -3k_B T$. This nonlinear behavior is typical for many HP sequences[33,38] and is caused by kinetic traps.

For the HP+ sequence in Figure 31B, more than 90% of the chains undergo fast Metropolis folding. Figure 31B shows features somewhat similar to a chevron plot, except that at high HH sticking (more negative ϵ) the Metropolis relaxation rate k_r approaches a constant instead of increasing monotonically with decreasing ϵ as in the chevron plot in Figure 19. This is because the rate of forming a favorable contact in Metropolis dynamics does not increase with the contact's favorability. Therefore, there is an upper limit to downhill folding rates, the "diffusion limit" of the model. Figure 31B shows that the folding rate k_f increases as ϵ decreases when HH sticking is weak. In this range, increasing HH sticking leads to faster Metropolis folding because breaking of already formed favorable contacts during the stochastic folding process is increasingly discouraged. The folding rate levels off only when breaking favorable contacts becomes virtually impossible. Typical chevron-like behavior is absent for this sequence

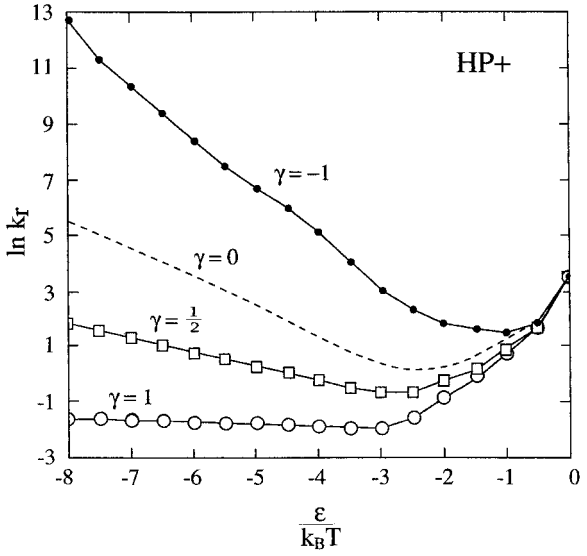

Fig. 32. Varying Kawasaki dynamics. Folding relaxation rates ln k_f vs. ϵ for the HP+ sequence in Figure 1 using generalized Kawasaki dynamics with different γ. $\gamma = 0$ is the same as Figure 19 in the text. Relaxations estimated for Δy from $\approx$ 0.65 to 0.4 with $\gamma = -1$, ½, and 1 are included for comparison.

because the Metropolis relaxation rate k_r is dominated by its unfolding rate for $\epsilon > -3k_BT$. For this sequence, the folding rate k_f saturates when the unfolding rate just begins to become less significant at $\epsilon \approx -3k_BT$. However, Metropolis dynamics can be compatible with chevron-like behavior for other sequences[182]: Figure 31B shows such behavior for a hypothetical energy landscape.

The slopes of model chevron plots depend on the model dynamics. Figure 32 shows chevron plots with different Kawasaki parameters in Eqs. (12) and (13). There are upper limits to downhill folding rates because of solvent viscosity[112,189–192] and intrinsic energy barriers to covalent bond rotations. This means that Kawasaki-like model dynamics may not apply when the sticking energy is too strong. At that point a leveling off of folding rate similar to that for Metropolis relaxation in Figure 31B is expected.

Finally, we comment on simulation standards. There are two problems that appear in some studies.[115] First, it is often only practical to compute first-passage times in some simulations (see, e.g., Refs. 32 and 33) rather than a more proper ensemble-averaged evaluation of the full kinetics. The first-passage time is the time required for a chain trajectory to first locate the native state. Second, another trick to avoid computational limitations is to simulate chains under denaturing conditions,[100,101] where searching is relatively fast on bumpy landscapes. This too helps computational speed, but it is not a realistic reflection on the true folding kinetics under native conditions.[115] Although our previous folding simulations[33,38] have been criticized as being slow, in particular for reporting folding times with more Metropolis time steps than the total number of conformations (40,617 for $n = 13$),[5,48,193,194] these criticisms are inappropriate. First, we do not consider it justified to compare model folding kinetics under strongly native conditions with first-passage time simulations under unfolding conditions. Under weak HH sticking conditions, the first-passage time can be much shorter than the folding time required to achieve a significant native population, because the chain would not remain in the native state. For example, despite our long folding times under strongly native conditions, the shortest mean first-passage time among the $n = 13$ HP sequences using Metropolis dynamics is only $\approx 3 \times 10^3$ time steps (Fig. 20b in Ref. 38). Second, under relatively strong folding conditions that guarantee $\approx$ 98% equilibrium native population for the $n = 13$ HP sequence in Figure 1, our Metropolis dynamics results indicate that half of the chain population is folded in $\approx 140A = 4{,}060$ Metropolis time steps (see the $\epsilon = -6k_BT$ relaxation curve in Fig. 30A, where A = 29). This example shows that for some short 2D HP sequences the number of time steps needed for folding can be significantly smaller than the total number of conformations.

APPENDIX C
MODELING THE TEMPERATURE DEPENDENCES OF RATE CONSTANTS

There are three levels of approximation for the physical forces that may determine the temperature dependence of the kinetics of folding and unfolding. At the first level of approximation, we can suppose that the HH interaction energy ϵ is independent of temperature. Then, the kinetics is determined by the balance between the HH sticking interactions and the conformational entropy of the chains. In general, a lowering of the temperature will speed up the folding on funnelscapes but also can deepen kinetic traps on bumpy landscapes. This level of approximation will not predict cold denaturation.

The second level of approximation recognizes that the hydrophobic interaction can itself depend on temperature, $\epsilon = \epsilon(T)$. The combination of these two ingredients are the basis for our Figure 22. This improves the predictions; it now predicts cold denaturation, but it still leaves some error in capturing the shape of the experimental curve for the unfolding rate vs. temperature and errs in predicting that two-state kinetics at relatively high temperatures will be dominated by entropy.

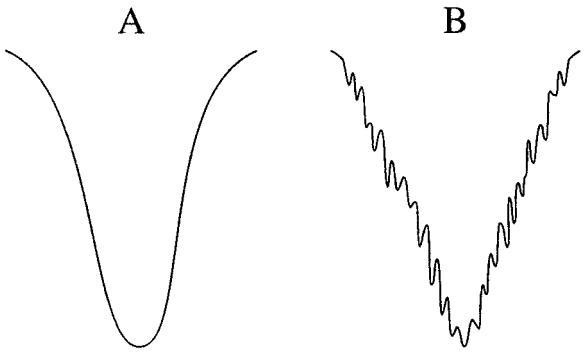

Fig. 33. Coarse-grained landscapes look smooth (**A**) even though bond rotation barriers lead to roughness (**B**) on a smaller scale.

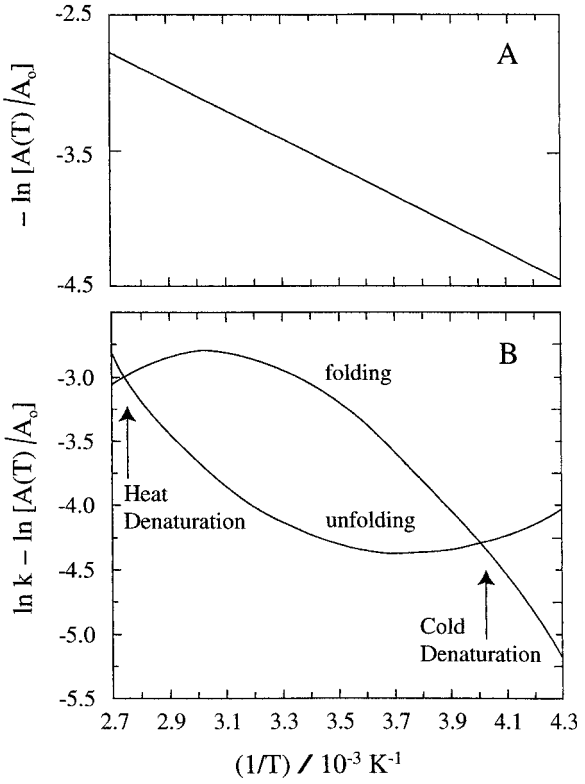

Fig. 34. If bond rotations have Arrhenius rates, $-\ln[A(T)/A_0]$ + constant $\sim 1/T$, as in (**A**), the model predicts folding and unfolding rate constants shown in (**B**), for the fast-folding HP+ sequence in Figure 1 (from Ref. 8).

The third level of approximation recognizes that bond rotations involve enthalpic barriers (see Fig. 33). To account for this, the intrinsic rate of conformational transition, A in Eqs. (11) and (12), will now become temperature dependent, $A = A(T)$. The timescale is defined by a reference temperature $T = T_0$ such that model time $t \equiv \tau/A_0$, where $A_0 \equiv A(T_0)$. A is a normalization constant for transition probability per unit time step [from τ to $\tau + 1$, see Eqs. (11) and (12)]. Therefore, physically, A should increase with decreasing temperature so that the intrinsic conformational transition rate is lower at lower temperatures. For instance, $A(T)$ may take the form

$$A(T) = A_0 \exp\left[\frac{\Delta G_b^{\ddagger}}{k_B}\left(\frac{1}{T} - \frac{1}{T_0}\right)\right] \qquad (14)$$

where $\Delta G_b^{\ddagger} > 0$ is a free-energy barrier to bond rotation. Now the temperature-dependent rate will be $k = k_0 A_0/A(T)$, where k_0 is the rate computed by using $A = A_0$. Combining a function $A(T)$ of this form with the model used for Figure 22 gives the results in Fig. 34, which provides a reasonable accounting for the experimental data of Chen et al.[70] (Fig. 35).

This treatment assumes that the temperature dependence of folding and unfolding rates of fast-folding, two-state proteins has its origin mainly in the temperature dependences of the intrachain contact interaction and bond rotation kinetics. This picture is consistent with recent experiments showing that the folding kinetics of Protein L becomes Arrhenius once the rates are corrected for the temperature dependence of protein stability.[169]

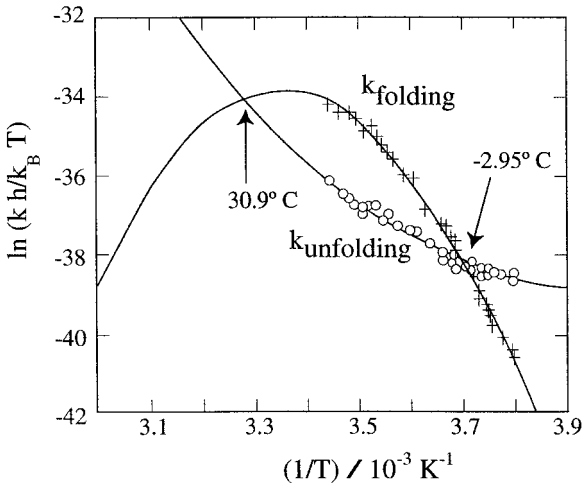

Fig. 35. Experimental temperature dependences of rates for T4 lysozyme (from Ref. 70).