Proc. Natl. Acad. Sci. USA
Vol. 94, pp. 6170–6175, June 1997
Biophysics

Folding funnels and energy landscapes of larger proteins within the capillarity approximation

PETER G. WOLYNES

School of Chemical Sciences, University of Illinois, Urbana, IL 61801; and Fogarty International Center, National Institutes of Health, Bethesda, MD 20892

Contributed by Peter G. Wolynes, April 1, 1997

ABSTRACT **The characterization of protein-folding kinetics with increasing chain length under various thermodynamic conditions is addressed using the capillarity picture in which distinct spatial regions of the protein are imagined to be folded or trapped and separated by interfaces. The quantitative capillarity theory is based on the nucleation theory of first-order transitions and the droplet analysis of glasses and random magnets. The concepts of folding funnels and rugged energy landscapes are shown to be applicable in the large size limit just as for smaller proteins. An ideal asymptotic free-energy profile as a function of a reaction coordinate measuring progress down the funnel is shown to be quite broad. This renders traditional transition state theory generally inapplicable but allows a diffusive picture with a transition-state region to be used. The analysis unifies several scaling arguments proposed earlier. The importance of fluctuational fine structure both to the free-energy profile and to the glassy dynamics is highlighted. The fluctuation effects lead to a very broad trapping-time distribution. Considerations necessary for understanding the crossover between the mean field and capillarity pictures of the energy landscapes are discussed. A variety of mechanisms that may roughen the interfaces and may lead to a complex structure of the transition-state ensemble are proposed.**

Proteins can be thought of as mesoscopic systems; that is, they are large in atomistic terms but are too small to be completely analyzed by macroscopic reasoning alone. The energy landscape theory of protein folding describes folding kinetics through a statistical characterization of the energies of different conformations (1, 2). This makes a bridge with the theory of phase transitions in disordered systems. The theory suggests that rapidly foldable proteins have an energy landscape dominated by a funnel to a large basin of attraction, the native state, and many small rugged features that give rise to trapping in local minima (3–5). Reliable folding requires the guiding forces of the funnel to be strong enough to overcome the entropy of the unfolded states. This must occur at a temperature such that trapping (also favored at low temperature) is not so strong that flow down the funnel is too slow. Much of the qualitative content of this picture has been confirmed through the computer simulation of small lattice models of proteins (5–8). Quantitatively, the analytical theories of folding landscapes have often made use of mean field descriptions that should be most accurate when the protein is small enough that we can think of each rearranging subunit of the chain as being able to interact with a fair fraction of the others at any time (1, 2, 9–13). This is true of the smallest lattice models. If the rearranging units are appropriately taken to be small segments of helices rather than individual residues, it is also true for the smallest globular proteins in nature, which have about 70 residues (4).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

This paper discusses how the folding dynamics of larger proteins can be described in terms of funnels and landscapes also in the opposite limit, where the protein is much larger than the range of the interresidue forces and the local correlations. In this limit we can define an interface or front between the folded and unfolded parts of a protein. Similar fronts, which are relatively free to move, exist between segments of the chain that are trapped in incommensurate misfolded configurations. To obtain a complete picture of the problem, these two aspects must be combined. The resulting capillarity model seeks to address the asymptotic behavior of folding kinetics with increasing protein size. As for the funneled aspects of the landscape, the capillarity picture is just like that of a first-order phase transition in a cluster. Regarding the rugged features of the landscape, the capillarity picture of the energy landscape is based on the droplet model of glasses, spin glasses, and random ferromagnets (14–19). For at least a decade, despite many simulation studies and much analysis, it has remained controversial at what size, if any, such disordered systems exhibit a crossover from the mean field behavior to the one envisioned by droplet arguments. Experiments on mesoscopic spin glasses show features in common with both limits (20). One apparent difference between the predictions of mean field and droplet theories is the way kinetics changes with size. But it is hard to convincingly debate this issue in the biophysical context because natural proteins, being the result of evolution, are not truly scaleable at the experimenter's discretion. Instead, large proteins typically have identifiable domains. Although the mean field theories of folding and trapping give barriers scaling linearly with chain length, these barriers still are predicted to be modest for proteins in the size range of a typical naturally occurring domain (11). Weaker scaling with chain length than mean field theory has already been suggested several times (21–23). Finkelstein and Badredtinov (21) obtain a barrier scaling like $N^{2/3}$, using the capillarity picture in an elegant way that, however, neglects ruggedness of the landscape. Thirumalai (22) focuses on the rugged features of the landscape and uses the scaling theory for glasses (18, 19) to obtain a barrier scaling as $N^{1/2}$. Simulations of proteins selected from a random pool to be fast folders or designed using the minimal frustration principle suggest a logarithmic scaling of the barrier height at the temperature of fastest folding but linear scaling with N of slow-folding events at low temperature (23). The general treatment here unifies these disparate arguments and observations by highlighting the importance of fluctuations and of how the barriers depend on thermodynamic state as well as chain length. The treatment also emphasizes that the dynamics in the capillarity picture is totally consistent with the funneled landscape pictures used phenomenologically.

The organization of this paper is as follows: In the first section the funneled aspects of the landscape are discussed within the capillarity approximation. The similarity with the kinetic description usually discussed via mean field theory is made clear. In the next section the glassy landscape ruggedness is discussed using an explicit capillarity-based picture generalizing Thirumalai's scaling arguments. These aspects are

combined to discuss the "typical" behavior of folding times for a large protein. The "fine structure" of the free-energy profile and other fluctuation effects are discussed in the following section. This analysis shows the delicacy required of computer simulations addressing size scaling issues and why fluctuation effects need to be considered in interpreting recent simulations. Further questions raised, especially involving the crossover between the mean field and capillarity approximations to the landscape, are discussed in the final section.

Capillarity Description of a Folding Funnel. Bryngelson and Wolynes (9) examined folding barriers with capillarity ideas in 1990. Following classical nucleation theory for first-order transitions, they wrote the free energy in terms of a progress coordinate, the number of residues folded, N_f. Their expression contains a linear "bulk" term and an interfacial term scaling like $N_f^{2/3}$:

$$F(N_f) = (f_F - f_u)N_f + \gamma N_f^{2/3} \qquad [1]$$

The bulk term depends on the free-energy difference per particle Δf between folded (f_F) and unfolded (f_u) protein; the small value of $\Delta f = f_F - f_u$ under folding conditions reflects the near cancellation of the entropy of unfolded state and the stabilizing energy of the native structure. The "interface" term γ was taken to be largely energetic. Putting in the typical stabilization of proteins under physiological conditions ($10k_BT$), they concluded that of the order $\approx$100 residues need to be ordered at the folding transition state, a number comparable to a protein domain size. At this level, that a biopolymer is being studied is largely immaterial and, in fact, nearly the same description was used somewhat earlier by Reiss *et al.* to describe mesoscopic cluster freezing transitions. In that context they made an observation that is important for proteins as well: the bulk transition temperature at which Δf vanishes does not coincide with the transition temperature of the cluster T_F. The temperature at which the two global free-energy minima have the same free energy is depressed by the surface contribution, as has long been known for the boiling of drops (25). Following Reiss *et al.* (24), we can then rewrite Eq. **1** by showing explicitly the temperature dependence of the free energy referenced to the transition temperature

$$F_{id}(\tilde{N}_f) = (-\tilde{\gamma} + \Delta H(T - T_F)/T_F)\tilde{N}_f + \tilde{\gamma}\tilde{N}_f^{2/3}, \qquad [2]$$

where we have written $\tilde{\gamma} = \gamma N^{2/3}$ *and scaled* N_f by the chain length N. At T_F, this is a crudely universal form for the "ideal" free-energy profile since $F_{id}(0)$ *must equal* $F_{id}(1)$. The temperature dependence of the stability depends on the enthalpy of unfolding ΔH. In this expression we see the normalized folded fraction $\tilde{N}_f = N_f/N$ can be used as a progress coordinate for the unfolding reaction here just as reaction coordinates highlighted in other theories of the funnel. The specific numeric coefficients above (which are used for concreteness in this paper) assume the protein or cluster is nearly spherical, so the curvature of the front is equally limited by all dimensions of the protein (see Fig. 1).

At T_F the free-energy profile is a rather broad curve as shown in Fig. 2. The maximum occurs at $N_f^{\neq} = {}^{8}\!/_{27}$ N. The barrier can be ascribed to the interface term and is given by $\Delta F^{\neq} = {}^{4}\!/_{27}\ \gamma N^{2/3}$. This is the barrier scaling obtained by Finkelstein and Badredtinov (21), using a more elaborate treatment of the interface contributions and a more careful treatment of the protein shape. These effects can be important; for example, if the protein is cylindrical, the front will orient orthogonal to the long axis and the barrier will not depend on the total cylinder length, a case reminiscent of coiled coils such as gcn4, where folding was recently studied (26).

The breadth of the free-energy profile has consequences for the kinetic description and is reflected in the size of the transition

FIG. 1. A schematic view, according to the capillarity picture of structure of a protein (lysozyme is shown for illustration) once it is advanced partially down the folding funnel. When several phases of the protein are possible, they can lead to a rather complex interface, as illustrated here. Thermal- and disorder-induced roughening of the interfaces can smooth out the transition region seriously. The figure shows several kinds of ordering that are possible: a region in which side chains are completely ordered (blue), a transition zone with topologically correct folding but no side-chain order (red), a molten-disordered phase (yellow), and the random coil (green). Having such a gradual progression of phases can lower the activation energy but broadens the interface between completely folded and unfolded regions. The front is shown, for clarity, as progressing from one end of the protein to the other. The most well ordered part may be buried in the core, depending on the relative surface tensions between the phases and the solvent.

state region δN_{TST}, defined as the range over which F(N) changes by k_BT. This range in the fractional progress coordinate is $\delta\tilde{N}_{TST} = \sqrt{k_BT/(\delta^2F/\delta^2\tilde{N}^2)}$. Using Eq. **2** at T_F, the number of residues displaced in moving over the transition region is approximately $\delta N_{TST} \cong 2.\sqrt{2}/3N^{2/3}\sqrt{k_BT/\gamma}$, showing the barrier becomes broader in terms of displaced residues with increasing chain length. An elementary move in folding dynamics is thought to involve displacing a loop (27) whose length scales like $N^{1/3}$. Thus, crossing this region will take many elementary moves and will be expected to be at least at the border of diffusive behavior. Even in the small-size regime, Socci *et al.* (28) have shown that, for the 27-mer lattice model, the diffusive dynamics works quite well. Thus, we see at T_F, the capillarity argument, like that of the mean field, suggests that $\tilde{N}_f$ can be taken as a reaction coordinate but must be treated as diffusive and traditional transition state theory should not be used. The only difference from the mean field funnel description is that capillarity theory assumes most contacts made in the partially folded protein are contiguous in physical space, whereas mean field estimates allow them to spread out. The capillarity model, therefore, obtains a different scaling of the thermodynamic barrier with chain length, but the phenomenological analysis is unchanged. It is worth noting that predictions of mean field barrier heights (11) at T_F

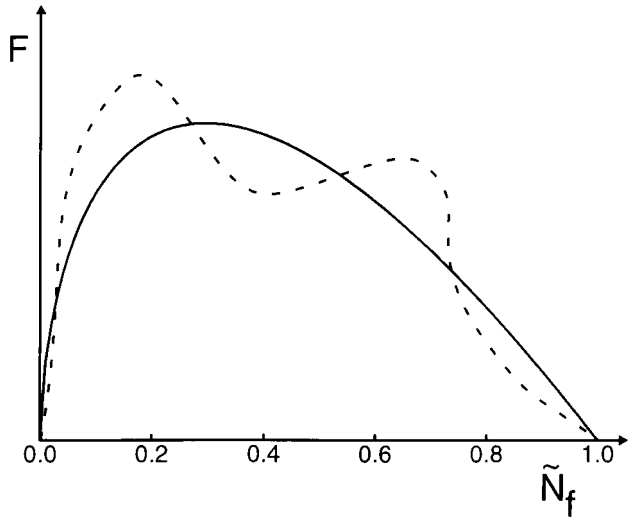

FIG. 2. The funnel free-energy profile in the capillarity approximation. The solid curve is the ideal profile of the free energy versus the folded fraction, assuming no heterogeneity of stabilization energy. An actual profile differs by a fluctuation that is random and Gaussianly distributed due to this heterogeneity. This is the "fine structure" of the folding funnel. A specific such profile is indicated by a dotted line. Notice the cusp-like minima occur at precisely $\tilde{N}_f = 0$ and $\tilde{N}_f = 1$, because the capillarity theory is an asymptotic result for large N_{TOT}. Simulations and mean field theory show smooth minima at nonextreme values of the order parameters.

actually are smaller than Finkelstein's (21) specific estimates for barriers even for $N \approx 100$, suggesting that the contiguous configurations assumed by capillarity need not dominate in this size range over a set of more diffuse arrangements of contacts with greater diversity. Eventually as temperature decreases when $T \ll T_F$, the curvature of the fronts will be large and the barrier to folding will be nearly independent of chain length. Polymeric effects of chain connectivity should give a weak logarithmic dependence on size for the purely thermodynamic barrier (22).

In using capillarity arguments we note there are two reasons proteins may have quite a complicated structure for the interface between "folded" and "unfolded" regions. First, thermal fluctuations and heterogeneity of native contacts roughen the interface. Indeed, because of the reduction in the number of neighbors when going from three dimensions to the two dimensions of the interface, the roughening temperature of the interface could be quite a bit lower than the folding transition (29). Another source of interfacial width is the possible existence of what can be thought of as multiple phases of the protein in bulk. Even when the equilibrium unfolded state is expanded, a metastable compact liquid or even liquid crystalline state, or one topologically correct without side chain ordering may exist and have been theoretically suggested. Being intermediate in structure, these phases normally will partially wet the interface between completely folded and unfolded states, as shown in Fig. 1. This reduces the interface energy γ and the thermodynamic activation barrier. Thus, for instance, although in the interior of the folding phase side chains may be ordered, they need not be ordered in the interface. If the interface is very rough and wide enough to be comparable to the diameter of the protein, the capillarity picture will be inappropriate, but one crosses over to a more mean field-like description with contacts spread everywhere.

Because of the diffusive nature of the barrier crossing at T_F, the time for folding will involve both the thermodynamic barrier $\Delta F^{\neq}$ and the reconfiguration times, which themselves can involve an activated barrier crossing. The description using diffusive theory (2, 8) gives a simple dependence on the typical reconfiguration time in the unfolded state, τ_0^{RC}:

$$\tau = \tau_0^{RC}(N,T)e^{\Delta F^{\neq}/k_BT}. \quad [3]$$

In the case where $T \ll T_F$, with its resulting "small" folding nucleus, the free-energy profile will be sharper at the transition-state region, allowing the barrier crossing to be less diffusive and more direct as in transition-state theory. If the inherent thermodynamic barrier is small, however, the subsequent growth of the folded region for a large enough protein can be rate limiting. The time for this thermodynamically downhill step will also be proportional to τ_0^{RC}. Within the capillarity picture, as in mean field theory, the reconfiguration times depend on escaping from traps and on the ruggedness of the landscape. These effects have been neglected in earlier work by focusing primarily on the case where the unfolded part of the chain is not compact (21). This requires T_F to exceed the random heteropolymer collapse transition, T_C, otherwise an additional barrier for expanding the chain, scaling with N, is introduced. T_C is *a fortiori* even higher than the glass transition temperature, so ruggedness effects can be neglected for such a well designed, minimally frustrated protein. It is not yet clear whether proteins are in fact frustrated so little. If the stability gap is not so large as to allow T_F to exceed T_C, the unfolded segment will be collapsed and possibly transiently misfolded. We must then consider trapping in determining the N and T dependence of the trap escape processes that enter τ_0^{RC}.

Capillarity Description of Reconfiguration Dynamics. The capillarity description of first-order phase transitions is well established by experiments documenting the reduction of freezing temperature with size. The size dependence of glassy phenomena, the sort needed to understand trapping, is hardly explored. In mesoscopic samples of magnetic spin glasses, elegant experiments of Weissman (20) show behavior reminiscent of mean field theories, although in some cases droplet-like excitations seem to be involved, too. Experiments on glass transitions of fluids confined to pores also sometimes show a reduction in T_G, but there are complications due to surface anchoring (30). Nevertheless, it is worth examining where theory leads.

The glass transition of random heteropolymers is believed to be a "random first-order transition" involving a discontinuous change in heat capacity and an entropy crisis similar to the supposed ideal glass transition of supercooled liquids (31, 32). The scaling theory of dynamics of liquids approaching such an ideal glass transition (18–19) has been formulated partially in terms of capillarity arguments. We can explicitly adapt those arguments to discuss the finite size effects appropriate to proteins. Above the thermodynamic transition a system that will undergo a random first-order transition can be thought of as a system that has an exponentially large number of metastable phases. In the heteropolymer context these are the trapping states. Above a dynamic transition temperature T_A, which is greater than T_G, each of those phases becomes even locally unstable to thermal motions. Above T_A, the chain dynamics is slowed down by caging but is much like a free chain. Below this temperature, on the other hand, when the interactions are short ranged, reconfigurational movements become slow and activated. The driving force for escape from one to another is the configurational entropy density. When the system is trapped in a given minimum it costs energy to move into another one, but there are so many ways to do this that escape becomes favorable. If a region of size N_e monomers to reconfigure the free-energy cost is

$$F_{trap}^{RC}(N_e) = -Ts_cN_e + f_I(N_e) \quad [4]$$

where s_c is the configurational entropy per residue and $f_I(N_e)$ is an interfacial energy term representing the mismatch be-

tween the configurations, each one of which is a local minimum, naively the interface energy f_I would scale as a surface term $N_e^{2/3}$. A detailed approximate calculation for the random heteropolymer yields this result (33) for the interface energy and an approximate quantification of trapping barriers. To understand the scaling we must recognize also that the existence of many "phases," which represent proteins in globally different local minima, however, reduces this surface energy by means of a wetting phenomenon such as the one we discussed for a finite number of phases at the end of section two of this paper.

As discussed earlier for specific, qualitatively different phases, a series of different phases also can decorate the interface, making it very broad and lowering the barrier. For highly curved interfaces f_I will behave like a surface term, but as the reconfiguring region increases in size, the effective surface tension will diminish through wetting as the interface straightens. A detailed argument much like the one for the random field Ising magnet (16, 18) finally gives a contribution that scales like $\alpha\Delta E N_e^{1/2}$, where ΔE is the root mean square of the energy variance per particle of the trapped states. In more elaborated theories of random systems there can be a more anomalous scaling such as $N_e^{1/2+x}$. The maximum of $F(N_e)$ gives the barrier for reconfiguration, $\Delta F_{trap}^{\neq RC}$ and we can write $\tau_0^{RC} = \tau_0 \exp \Delta F_{trap}^{RC}/k_BT$, as the typical reconfiguration time. For an infinite system the critical size corresponding to the maximum is $N_e^{\neq} = (\alpha\Delta E/Ts_c)^2$ and the barrier is $\Delta F_{trap}^{\neq RC} = \alpha^2\Delta E^2/2Ts_c$. The barrier is the same as the empirical Vogel–Fulcher law and depends on the configurational entropy density just as in the Adam–Gibbs description of glassy dynamics (31), but the number of units in the reconfiguring region is quite a bit larger than the Adam–Gibbs result for the critical size that scales only like the s_c^{-1} *rather than* s_c^{-2} predicted by random first-order transition scaling, where the so-called correlation length exponent is $\nu = 2/d$.

The capillarity argument shows that well *above* T_G the reconfiguration barriers are finite (i.e., they do not scale with chain length) just as the thermodynamic barriers to folding are chain length-independent far *below* T_F. Nevertheless there is a connection of the capillarity result to the famous Levinthal estimate, which arises in the mean field theory. We can see that at any temperature, the reconfiguration barrier is $\Delta F_{trap}^{\neq RC} = \frac{1}{2}s_c N_e^{\neq} = \frac{1}{2}S(N_e^{\neq})$. This is one-half of a renormalized Levinthal entropy for the critical sized drop $S(N_e^{\neq})$. The barrier now depends on the actual configurational entropy of the critical drop, not the infinite temperature value of the entropy of the entire system S_o. The barrier thus grows as T_G approaches, but more slowly than in the mean field theory, because the random wetting phenomenon allows the entropy to be gained in stages. At T_G the energy landscape for an infinite random heteropolymer is self-similar for different size regions. There are minifunnels within minifunnels.

Once $N_e^{\neq}$ grows to a size such that the rearranging region is comparable to the protein diameter, the interface cannot get any larger. Within the capillarity approximation, the driving force term can now be neglected, and again the barrier involves moving a front between different trap states across the molecule, giving a barrier $\Delta F_{trap}^{\neq RC} = \alpha\Delta E N^{1/2}$. This is the result obtained more elliptically by Thirumalai (22). We now see it is an asymptotic result valid near T_G.

Reconfiguration dynamics then gives a contribution to the apparent folding barrier that is independent of N at high T and scales like $N^{1/2}$ for T near T_G.

The same wetting phenomenon operating for traps can change the scaling of the thermodynamic barrier, too. If the randomness is large enough, the native state is hardly distinguishable from any other trap in energetics. The resulting multiple-phase wetting resembles disorder-induced roughening of the interface.

When we combine the results for the thermodynamic barriers and the typical configurational diffusion barriers in the diffusive barrier-crossing expression, we obtain various different scalings of folding time for a "typical" protein with chain length under different thermodynamic conditions, as shown in Fig. 3. For very well designed proteins that can be made to stably fold well below T_F but above T_G, the asymptotic scaling of folding time is polynomial in chain length. There is no contradiction with well known NP completeness theorems (34) about predicting global minima of a random heteropolymer because well designed proteins are selected, minimally frustrated systems. If still above T_G but near T_F, the scaling for the time grows exponentially in size but the barrier rises sublinearly as $N^{2/3}$ in Finkelstein's (21) calculation. Near to T_G, the scaling even for a well designed protein *again* would be exponential but with a different power, $N^{1/2}$, for the barrier, as Thirumalai (22) suggests. This is also the behavior of the typical folding time for an unselected random heteropolymer.

Fine Structure of the Funnel Free-Energy Profile and of Reconfiguration Dynamics. The simplest phenomenological funnel description of folding either in mean field or capillarity approximations maps folding kinetics near T_F onto the diffusion of a one-dimensional progress coordinate. More coordinates are needed if different phases are taken into account (e.g., secondary structure formed versus disordered, collapsed versus noncollapsed, etc.), but another effect is also important. At any given value of the progress coordinate, in a specific protein, different contacts will be made to varying extents, depending on their energetics. Although the ideal funnel has a fairly predictable shape in either approximation, the heterogeneity of contact energies and local propensities modifies the free-energy profile for any specific protein, as shown in Fig. 2. This heterogeneity can be probed experimentally by protein engineering (35) and by elegant NMR experiments (B. A. Shulman, P. S. Kim, C. M. Dobson, and C. Redfield, personal

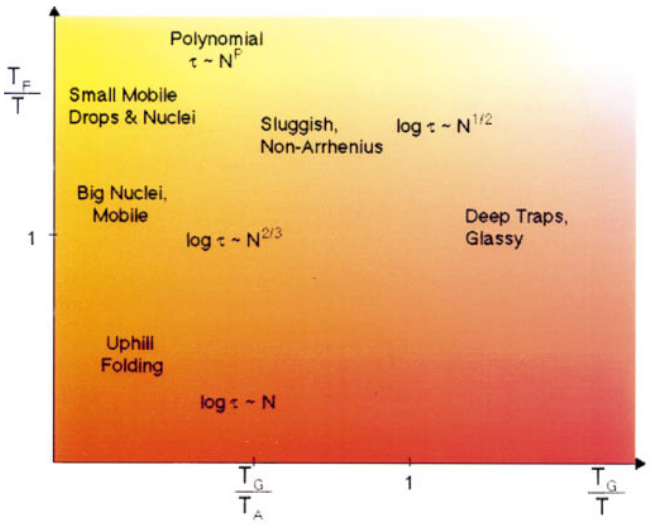

FIG. 3. The various regimes of scaling behavior for the typical folding time with length are indicated for different values of T_F/T, the folding temperature-to-ambient ratio, and T_G/T, the glass temperature-to-ambient ratio. At high temperatures, folding is always thermodynamic uphill and the barrier grows like N. At low temperatures, the rate is trap-dominated, and the typical barrier behaves like $N^{1/2}$, whereas the slowest escape (not indicated in the figure) will have bigger barriers, scaling like N. Fastest folding occurs for $T \ll T_F$ but $T \gg T_G$ where polynomial behavior of the folding time in N is to be found. This occurs for "well designed" proteins. The figure also shows the boundary between trap escape and quasi-free chain dynamics, as indicated by the mean field dynamic transition temperature T_A. A realistic value of T_F/T_G is 1.6. In this figure we assume temperatures are all less than the collapse temperature T_C. If T is bigger than T_C, ruggedness effects can be neglected.

communication). For specific proteins the resulting structural correlations can be discussed using free-energy functionals like those used for liquids or random magnets (37).

We now examine what the capillarity arguments give for the fluctuation effects on folding and trapping free-energy profiles. Let us assume the fluctuations in stabilization energy for different residues are small and independent, with a typical value being Δe_o. We expect Δe_o to be somewhat smaller than the ruggedness energy scale for compact states, ΔE. The ratio $\Delta e_o/\Delta E$ depends on the flexibility of the sequence code used for encoding the protein. At T_F, if the interface is sharp, a specific protein free-energy profile will differ from the ideal profile by a random contribution $\delta F_{rand}(\tilde{N}_f)$ so that $F(\tilde{N}_f) = F_{is}(\tilde{N}_f) + \delta F_{rand}(\tilde{N}_f)$. $\delta F_{rand}(\tilde{N}_f)$ will be characteristic of a diffusion process in N_f that must return at the end of its journey to the origin while never crossing the origin, if folding a single domain. Thus, $\delta F(\tilde{N}_f)$ is typically $\Delta e_o N^{1/2}(\tilde{N}_f(1-\tilde{N}_f))^{1/4}$. At T_F the transition state of the ideal funnel is nearly midway. Thus, the fluctuation in barrier height is $\sim \frac{1}{2}\Delta e_o N^{1/2}$. The thermodynamic activation barrier will be approximately Gaussianly distributed with this width. Taking the rather large value of $\Delta e_o \approx 1 k_B T$, one finds fluctuations as big as $5 k_B T$ for a 100-mer. In fact, if $\delta\Delta F^{\neq}$ is sufficiently large to overcome $\Delta F^{\neq}_{ideal}$, the folding event will be broken up into two parts (i.e., a kinetic intermediate will exist). The probability of finding a sequence that folds 100 times faster than typical is about one-third for such a large Δe_o. It becomes necessary in computer simulations to examine the distribution of measured rates rather than, for example, selecting (after the fact) the fastest folders, if the model is to be compared with analytical work.

There are many different trapping configurations; simply knowing their typical behavior is not enough. Fine structure effects on the distribution of trapping escape times are predicted to be especially large (and more subtle than on the thermodynamic folding profile) if the capillarity argument is followed. Adding a Gaussian random contribution to Eq. **4**, which represents fluctuations in energy of the trapped states contacts just as was done for the folding profile, the typical trap escape free-energy profile becomes widely distributed because now the renormalized interfacial energy term scales in exactly the same way with chain length as the randomness. The mean and variance of the escape barriers are comparable. Equivalently said, the assumed scaling exponent, $x = 0$, is marginal for a random system's transition (39).† Although escape from the typical and, more importantly, the deepest traps will significantly slow as T_G is approached, there is now a significant chance that trapping can be avoided altogether through motions involving one of these low barrier traps, even for long chains. The wide distribution suggests that trapping dynamics will be strongly multiexponential as T_G is approached. Even when the overall thermodynamic barrier still is large enough to give single exponential kinetics, the widely distributed trapping times can modify the transmission coefficient computed using the pure diffusion description of the barrier crossing; instead, a frequency-dependent Kramers theory should be used. This acts to diminish the influence of trapping on the rate, since the weaker traps will be used in the crossing. At low temperatures, once the thermodynamic barrier is small ("downhill kinetics"), the entire folding kinetics will be nonexponential and reflect the trap distribution directly. As in the kinetic partitioning phenomenology (40, 41), a small fraction of protein molecules can fold on a fast track, while the bulk of them will be trapped in misfolded states.

Although the heterogeneity effects on trapping can allow a small fraction of fast track folders, they also lead to a finite fraction of very slow folding molecules. The capillarity picture yields a roughly Gaussian form (except near $\Delta F = 0$) for the distribution of escape barriers

$$P(\Delta F^{\neq}) = \frac{1}{\sqrt{2\pi N \Delta E^2 \lambda}} \exp - \left(\frac{\Delta F^{\neq} - \Delta F^{\neq RC}_{Trap}}{2N\lambda\Delta E^2} \right)^2 \quad [5]$$

so the distribution of long escape times is roughly log normal. At T_G, the most probable escape time scales as $N^{1/2}$ as discussed in the previous section, but the mean folding time is dominated by the deepest traps. Averaging the escape time, which depends exponentially on the barrier over the log normal distribution, the apparent activation energy for the slowest events then scales linearly in N, just as it does from mean field theory. Bigger molecules can sample less and less typical, deeper traps. The linear dependence of the longest escape time was noted in recent simulations and may also be due to topological constraints on motion (23). This wide variance of escape times also was predicted at all temperatures on the basis of the random energy model (2). It is interesting that it persists below T_G in the capillarity description as well. The effects of native contact heterogeneity near T_F and trapping due to glassy dynamics near T_G are very similar and scale in the same way with N. The fluctuations away from the ideal funnel free-energy profile will typically act to slow the folding from the ideal profile result, because it is the big barriers that count for a long, random walk in a one-dimensional random potential (42). Below T_G, folding times, even of well designed proteins, will be quite sensitive to sequence.

Both the fluctuational fine structure of the free-energy profile and the broad distribution of trapping times from the glassy dynamics suggest the largest proteins will fold near T_G by a highly intermittent, partially progressive escape from traps, much like a Levy flight (42). This may be studied using single-molecule detection techniques.

Discussion. The analysis presented here shows that the energy landscape description of protein folding has much the same mathematical structure within the capillarity approximation as when mean field theories are used. Just as in mean field approaches, a folding funnel with a diffusive progress coordinate can be defined. Folding kinetics reflects both the free-energy profile of this coordinate and the reconfigurational motions that depend on the nature of glassy dynamics representing escape from local traps. The questions arise then of what determines which limit, mean field or capillarity based, is more appropriate for describing real proteins, and how can we find this out?

The crossover between mean field-like and capillarity behavior for the funnel characteristics depends on the breadth of the interface compared with the protein size. Even in the simplest analyses, this breadth will be proportional to the range of the interactions. For hydrophobic forces this range may be as small as a water molecule's diameter or as large as a residue's diameter. It still is longer for electrostatic forces. Using the residue size range, much of a small protein ($\approx$60 residues) is at the surface, arguing for the mean field picture. The breadth of the interface also increases with the heterogeneity of contact energies, eventually leading to disorder-induced roughening of the interface. There is also the possibility of interpolated, partially ordered phases (e.g., molten globule) broadening the interface. The magnitude of these effects is hard to estimate currently. One difficulty is that database contact potentials still are rather too crude to be sure of the size of heterogeneity effects. Also, the energetics of

†For concreteness, the exponent phenomenology for a random first-order glass transition has been followed in this paper. More involved phenomenologies in which the interface energy, randomness contributions, and their distributions and even activation barriers have anomalous exponents for size scaling are possible (see refs. 15 and 16). These possibilities that allow the glass state to be a "chaotic" phase (39) would give rise to considerations similar to those here but with quantitative modification.

detailed side-chain packing, which determines whether a protein "sublimes" or "melts," remain uncertain.

Experimentally, recent studies using NMR of the progressive denaturation of lactalbumin (≈140 residues) by urea do seem to indicate the disordering of the molecule by means of a front moving across the structure (B. A. Shulman, P. S. Kim, C. M. Dobson, and C. Redfield, personal communication). This suggests there is a crossover to capillarity behavior, but greater precision will be needed to pin down the interface width. According to the one-dimensional funnel picture, the equilibrium front should be closely tied to the landscape relevant to kinetics, but if additional order parameters are involved, this is less clear. The existence of a sharp, contiguous front could be probed kinetically by protein engineering using simultaneous mutations of contiguous sites, which could affect the fine structure of the funnel.

The existence of fine structure fluctuations in the capillarity picture has an impact on understanding the constraints faced in the evolution of protein-folding energy landscapes. Although the minimal frustration principle is still needed for reliable folding without traps, the fine structure effects on trapping apparently may allow a small fraction of the population of a molecule with a long, random sequence to fold quickly, even when the bulk of the population does not. Thus, the minimal frustration constraint is more appropriately thought of as one on yield rather than speed per se. Conversely, to achieve very fast folding may require, in addition to the minimal frustration constraint, that very stable contacts be judiciously placed in the structure so that even the ideal capillarity barrier is reduced. Such a feature may be correlated with the existence of foldons or modules in the larger proteins (36).

I thank C. Dobson, W. Eaton, M. Oliveberg, J. Onuchic, S. Plotkin, A. Szabo, and J. Wang for enjoyable discussions, and Ben Shoemaker for his assistance with the figures. The work at the University of Illinois was supported by National Institutes of Health Grant PHS R01 GM44557. This paper was written while I was a Scholar-in-Residence at the Fogarty International Center for Advanced Study in the Health Sciences of the National Institutes of Health (Bethesda, MD).

1. Bryngelson, J. D. & Wolynes, P. G. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 7524–7528.
2. Bryngelson, J. D. & Wolynes, P. G. (1989) *J. Phys. Chem.* **93,** 6902–6915.
3. Leopold, P. E., Montal, M. & Onuchic, J. N. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 8721–8725.
4. Onuchic, J. N., Wolynes, P. G., Luthey-Schulten, Z. & Socci, N. D. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 3626–3630.
5. Bryngelson, J. D., Onuchic, J. N., Socci, N. D. & Wolynes, P. G. (1995) *Proteins* **21,** 167–195.
6. Dill, K. A., Bromberg, S., Yue, K., Fiebig, K. M., Yee, D. P., Thomas, P. D. & Chan, H. S. (1995) *Protein Science* **4,** 561–601.
7. Mirny, L. A., Abkevich, V. & Shakhnovich, E. I. (1996) *Folding Des.* **1,** 103–116.
8. Socci, N. D. & Onuchic, J. N. (1995) *J. Chem. Phys.* **103,** 4732–4744.
9. Bryngelson, J. D. & Wolynes, P. G. (1990) *Biopolymers* **30,** 177–188.
10. Plotkin, S. S., Wang, J. & Wolynes, P. G. (1996) *Phys. Rev. E* **53,** 6271–6296.
11. Plotkin, S. S., Wang, J. & Wolynes, P. G. (1997) *J. Chem. Phys.* **106,** 2932–2948.
12. Plotkin, S. S., Wang, J. & Wolynes, P. G. (1997) *J. Physique I* **7,** 395–421.
13. Takada, S. & Wolynes, P. G. (1997) *Phys Rev. E* **55,** 4562–4577.
14. McMillan, W. L. (1985) *Phys. Rev. B* **31,** 340–341.
15. Fisher, D. S. & Huse, D. A. (1988) *Phys. Rev. B* **38,** 373–385.
16. Villain, J. (1985) *J. Phys. (Orsay, Fr.)* **46,** 1843–1852.
17. Kirkpatrick, T. R. & Wolynes, P. G. (1987) *Phys. Rev. B* **36,** 8552–8564.
18. Kirkpatrick, T. R., Thirumalai, D. & Wolynes, P. G. (1989) *Phys. Rev. A* **40,** 1045–1054.
19. Wolynes, P. G. (1988) in *Proceedings International Symposium on Frontiers in Science*, eds. Chan, S. & DeBrunner, P. G. (Hans Frauenfelder Festschrift) (Am. Inst. Physics, New York), pp. 39–65.
20. Weissman, M. (1988) *Rev. Mod. Phys.* **60,** 537.
21. Finkelstein, A. V. & Badredtinov, A. Y. (1997) *Fold. Des.* **2,** 115–121.
22. Thirumalai, D. (1995) *J. Phys. (Orsay, Fr.)* **15,** 1457–1467.
23. Gutin, A. M., Abkevich, V. & Shakhnovich, E. I. (1996) *Phys. Rev. Lett.* **77,** 5433–5436.
24. Reiss, H., Mirabel, P. & Whetten, R. (1988) *J. Phys. Chem.* **92,** 7241–7250.
25. Rowlinson, J. T. & Widom, B. (1982) *Molecular Theory of Capillarity* (Clarendon Press, Oxford).
26. Sosnick, T., Jackson, S., Wilk, R. R., Englander, S. W. & De Grado, W. F. (1996) *Proteins Struct. Funct. Genet.* **24,** 427–432.
27. Hagen, S. J., Hofrichter, J., Szabo, A. & Eaton, W. A. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 11615–11617.
28. Socci, N., Onuchic, J. & Wolynes, P. G. (1996) *J. Chem. Phys.* **104,** 5860–5868.
29. Weeks, J. D., Gilmer, G. H. & Leary, H. J. (1973) *Phys. Rev. Lett.* **39,** 549–553.
30. Zhang, J., Liu, G. & Jonas, J. (1992) *J. Phys. Chem.* **96,** 3478–3480.
31. Kauzmann, W. (1948) *Chem. Rev.* **43,** 219–256.
32. Adam, G. & Gibbs, J. H.(1965) *J. Chem. Phys.* **43,** 139–146.
33. Takada, S. & Wolynes, P. G. (1996) *Prog. Theor. Phys. Suppl.*, in press.
34. Ngo, J. T. & Marks, J. (1984) *Protein Eng.* **5,** 313–321.
35. Fersht, A. R. (1997) *Curr. Opin. Struct. Biol.* **7,** 3–9.
36. Panchenko, A. R., Luthey-Schulten, Z. & Wolynes, P. G. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 2008–2013.
37. Shoemaker, B. J., Wang, J. & Wolynes, P. G. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 777–782.
38. Chayes, J. T., Chayes, L., Fisher, D. S. & Spencer, T. (1986) *Phys. Rev. Lett.* **57,** 1999–2002.
39. Bray, A. J. & Moore, M. A. (1987) *Phys. Rev. Lett.* **58,** 57–61.
40. Thirumalai, D. & Woodson, S. A. (1996) *Acc. Chem. Res.* **29,** 433–439.
41. Kiefhaber, T. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 9029–9033.
42. Bouchaud, J. & Georges, A. (1990) *Phys. Rep.* 127–293.