Buffed energy landscapes: Another solution to the kinetic paradoxes of protein folding

Steven S. Plotkin[†‡] and Peter G. Wolynes[‡§]

[†]Department of Physics and Astronomy, University of British Columbia, Vancouver, BC, Canada V6T 1Z1; and [§]Department of Chemistry and Biochemistry and Center for Theoretical Biological Physics, University of California at San Diego, La Jolla, CA 92093

The energy landscapes of proteins have evolved to be different from most random heteropolymers. Many studies have concluded that evolutionary selection for rapid and reliable folding to a given structure that is stable at biological temperatures leads to energy landscapes having a single dominant basin and an overall funnel topography. We show here that, although such a landscape topography is indeed a sufficient condition for folding, another possibility also exists, giving a previously undescribed class of foldable sequences. These sequences have landscapes that are only weakly funneled in the conventional thermodynamic sense but have unusually low kinetic barriers for reconfigurational motion. Traps have been specifically removed by selection. Here we examine the possibility of folding on these "buffed" landscapes by mapping the determination of statistics of pathways for the heterogeneous nucleation processes involved in escaping from traps to the solution of an imaginary time Schroedinger equation. This equation is solved analytically in adiabatic and "soft-wall" approximations, and numerical results are shown for the general case. The fraction of funneled vs. buffed proteins in sequence space is estimated, suggesting the statistical dominance of the funneling mechanism for achieving foldability.

The mechanisms of protein folding are manifold. This diversity can only be captured by a global statistical survey of the protein's energy landscape. What can be said about this landscape merely from the knowledge that proteins fold?

A foldable protein is a heteropolymer that through Brownian motion finds a particular structure within a short time and, once this structure is found, essentially stays in that structure. This requirement is both thermodynamic and kinetic. Thermodynamically, the folded state of a protein must be stable and tolerably unique. Although conformational substrates can be functionally important and are decidedly present (1), the specificity of proteins required for their work in the cellular network leads to strong constraints on the structure of the active sites and binding sites of proteins.

For random heteropolymers, thermodynamic foldability depends on the relation between the physiological temperature and the glass temperature, which depends on the amino acid composition. For some compositions, e.g., near homopolymers, T_G is low, whereas for other compositions using many chemically distinct amino acids, T_G could be higher than physiological temperatures. If physiological temperatures are very high, only a small fraction of all sequences will be foldable in a thermodynamic sense, because few sequences would have sufficient stability in the ground-state conformation against entropic costs. At sufficiently low temperatures, all sequences would have stable ground states, providing the solvent is such that interaction-free energies do not change much as temperature is lowered. But in this low-temperature case, the kinetics of folding would become the controlling factor. For a typical random sequence of amino acids, while the lowest energy state becomes thermodynamically stable only below the glass transition temperature T_G, at these low temperatures several other traps will also become stable. The average escape time from these traps scales exponentially in chain length. Thus for long enough chains folding will be slow.

One resolution of this conflict between kinetic and thermodynamic constraints is for proteins to have evolved to have compositions with $T_G < T$ but to be unusual sequences that nevertheless are stable in some configuration above T_G via the introduction of an energy gap in the density of states such that one energetic basin dominates. The energy landscape is funneled, with a single dominant basin (2). According to this view, proteins are atypical heteropolymers by having their folding transition temperature T_F above the glass temperature T_G to be expected ordinarily from the overall amino acid composition. This is the quantitative form of the principle of minimal frustration that was put forward by Bryngelson and Wolynes (3).

An extreme limit of the principle of minimal frustration would be for there to exist perfect consistency of interactions in the native state, an idea due to Gō and coworkers (4). In the space of all sequences, such perfection is less likely than merely achieving a satisfactory level of resolved frustration. Most microscopic potential models suggest that natural proteins still exhibit frustration in their ground state. The existence of this frustration is buttressed by the observation that many proteins can be stabilized further by single site mutations. Nevertheless, the perfect funnel model does describe the kinetics of folding of many proteins as a first approximation, as seen by the good agreement of ϕ values predicted by using Gō models with experiments (refs. 5–7 and references therein).

The iconoclasts among us, nevertheless, must ask whether the achievement of a funneled landscape with $T_F/T_G > 1$ is the only way the kinetic/thermodynamic requirements for folding can be met. Rather than finding sequences that increase the native stability under conditions where dynamics of a random sequence would be facile, we can imagine designing an abnormally trap-free landscape giving unusually high mobility throughout the misfolded parts of the energy landscape.

In this case, heteropolymers having compositions giving $T_G > T$ would easily have a ground state that is stable, but it would be necessary to design out the barriers for escaping the deep traps such that those barriers were atypically small. The density of low-energy states would not be atypical in this case; rather, what is atypical is the size of the kinetic barriers between the states. Such an energy landscape could be said to be "buffed." Even on such a buffed energy landscape, the native state still must have some extra stability to avoid the time-scale problems of a random search. But the extra stability in this scenario need not scale extensively with system size and may scale as a nonextensive fluctuation $\sim N^{1/2}$ or smaller.

It is intriguing that some of the successes of theories based on the strongly funneled landscape would be preserved for theories based on buffed landscapes. For example, because the ruggedness *en route* to the native structure is reduced and the effects of transient trapping diminished, native structure formation (as measured through ϕ values) will once again be dominated by polymeric properties. However, there should be observable differences between buffed and funneled landscapes in stability and robustness. The native state on a funneled landscape is marginally stable with respect to a large ensemble of unfolded states having significant

PHYSICS

BIOPHYSICS

entropy. The native state on a buffed landscape is marginally stable with respect to a few dissimilar low-energy states that are kinetically connected to it. Thus the thermal unfolding transition on a buffed landscape should be less cooperative than on a funneled landscape. The folded structure and folding mechanism are also less robust on a buffed landscape, because mutations can more easily open up or close off regions of searchable phase space. These differences between the properties expected for buffed sequences and natural proteins suggest that natural protein landscapes mainly are funneled rather than buffed.

Yet given the possibility of having a buffed energy landscape, the explanation for the prevalence of funneled landscapes must be sought in statistics and population biology (8). The question we address is whether the fraction of sequences with buffed landscapes is comparable or larger than the fraction of sequences with minimally frustrated landscapes in the conventional sense. This is the task of the present article.

Nucleation

Both folding and trap escape can be thought of as the growth of a new phase into a preexisting phase. This growth requires overcoming a barrier of finite height. According to classical nucleation theory, we can introduce a reaction coordinate N_e describing the amount of new phase, $0 \leq N_e \leq N$, where N is the length of the chain. The typical free-energy profile as a function of N_e is

$$\bar{F}(N_e) = F(0) - fN_e + \sigma N_e^z. \quad [1]$$

Here f is the bulk free-energy gain per residue in the new phase, and σ is the surface free-energy cost per residue. These parameters are each a function of temperature.

For a uniformly stable phase, the exponent $z = 2/3$. However, in a sufficiently heterogeneous medium the interface between phases can relax to a lower free-energy configuration that is roughened. The larger the nucleus, the more the interface can relax and the smaller the surface tension. For a wide class of models in three dimensions in this regime (9, 10), $\sigma = \sigma_o/N_e^{1/6}$, and the surface free-energy cost scales as $N_e^{1/2}$ rather than $N_e^{2/3}$. This is also true for a random first-order glass transition.

Setting $\partial \bar{F}(N_e)/\partial N_e = 0$ gives the typical critical nucleus size, $\bar{N}_e^{\neq} \equiv N\bar{n}_F^{\neq} = (z\sigma/f)^{1/(1-z)}$, where $0 \leq n_F \leq 1$. The typical free-energy barrier height is

$$\bar{F}^{\neq} = \sigma(1-z)(z\sigma/f)^{z/(1-z)} = \sigma(1-z)n_F^{\neq z}N^z. \quad [2]$$

Thus the barrier arises from surface cost and scales as N^z.

Generally for glassy systems at high enough temperature, trap escape becomes a downhill process, corresponding to a vanishing of the surface tension in the nucleation picture.

Atypically Rugged and Funneled Sequences

Suppose each amino acid was chosen from a pool with probability p_i^o for residue type i, where $1 < i < m$ with $m = 20$ and $\Sigma_i p_i^o = 1$. Shorter sequences may have compositions with probabilities of occurrence $\{p_i\}$ differing from $\{p_i^o\}$, but as the chain length increases, p_i approaches p_i^o. The probability of a composition $\{n_i\}$, $0 \leq n_i \leq N$, chosen randomly from a pool with natural abundances $\{p_i^o\}$ is given by a multinomial distribution. We are interested in the likelihood of finding a sequence with an unusually rugged landscape, thus we seek the value of the fluctuations $\delta(b^2)^2$ in the energetic variance per interaction b^2 as various sequences of length N are sampled. The conformationally averaged variance per contact b^2 for a given sequence is given by $\Sigma_{i,j=1}^m \delta\varepsilon_{ij}^2 p_i p_j$ (see e.g. ref. 11), where $\delta\varepsilon_{ij}^2$ is the variance of an element of the pair-interaction matrix between residue types i and j, e.g. the Miyazawa–Jernigan matrix (12).

The average sequence-to-sequence fluctuation in b^2, $\langle\delta(b^2)^2\rangle$, is obtained by calculating the fluctuations in the incorporation probabilities p_i, which decrease with increasing chain length as N^{-1}. The result is

$$\langle\delta(b^2)^2\rangle = \frac{4}{N}\sum_{i,j,k=1}^m \delta\varepsilon_{ij}^2\delta\varepsilon_{ik}^2 p_i^o(1-p_i^o)p_j^o p_k^o \equiv \frac{\sigma_b^2}{N}. \quad [3]$$

As chain length becomes large, the probability of variance b^2, $P(b^2)$, becomes an increasingly sharply peaked function about the average $\langle b^2 \rangle$. Applying the central limit theorem, the probability P_R that a sequence has ruggedness greater than say b_G^2 is just the integral of the tail of a Gaussian, i.e. an error function. Typically the interaction matrix and pool compositions are such that the thermal energy of the heteropolymer is well above the ground-state energy, i.e. $\langle b^2 \rangle$ is sufficiently small that the temperature $T > T_G$, where T_G is the glass temperature in the random-energy model (13). Later we will be interested in special sequences that are sufficiently rugged that without buffing they would be kinetically trapped and thermodynamically stable in low-energy states at biological temperatures. Because of the N dependence in Eq. 3, along with the fact that an error function has an exponential tail, such extra rugged sequences become exponentially rare with increasing system size (see Eq. 15).

Given an ensemble of sequences with ruggedness b^2, we now seek the fraction of these sequences with an energy gap larger than Δ. The relevant gap Δ is between the energy of the ground-state conformation and the energy of the next-lowest globally dissimilar structure. Hence the distributions of state energies $P(E)$ can be well approximated by Gaussian random variables of variance Ncb^2, where c is the number of interactions per residue. If there are $\Omega = e^{Ns_o}$ total dissimilar conformations (all assumed compact with the same c, here s_o is the conformational entropy per residue), $\Omega - 1$ of them must be Δ or higher in energy than a particular one. The particular ground state can have any energy but is most likely to have energy near the random-energy model ground state for a collection of Ω states (13). Any of the Ω states are candidates to be the ground-state structure. Then the fraction of sequences with a gap larger than Δ is given by the expression

$$p(\Delta|b) = \Omega\int_{-\infty}^{\infty}dE\,P(E)\left(\int_{E+\Delta}^{\infty}dE'\,P(E')\right)^{\Omega-1} \approx e^{-\sqrt{\frac{2s_o}{c}}\frac{\Delta}{b}}, \quad [4]$$

where the saddle point solution has been taken to obtain the last expression, as worked out previously in ref. 14. Note that the smoother the landscape (smaller b), the smaller the fraction of sequences gapped to Δ. We assume here no selection in ruggedness for funneled sequences, i.e. funneled sequences have the typical variance $\langle b^2 \rangle$ but have an atypical density of states.

Nucleation for Trap Escape

The alternative of buffed energy landscapes is not allowed in a strictly mean-field picture of folding dynamics, where the landscape is determined by the global statistical properties of the mean-field solution. However, for larger proteins, folding and trap escape depend on the spatially local (and/or local in sequence) properties of the energy landscape (15). For folding, these capillarity models resemble those of an ordinary first-order phase transition. On the other hand, in capillarity models of the glassy dynamics for trap escape based on the theory of random first-order phase transitions, fluctuations in the height of escape barriers scale in the same way as the typical barrier height (10). This motivates an estimate for the size of the ensemble of sequences with anomalously small escape barriers.

For a given sequence, the density of states is parameterized by two important energy scales, the variance b^2 and the energy gap Δ.[¶] There are other features of the energy landscape, however, not completely captured by the density of states such as the barrier height $F^{\neq}$ between dissimilar configurations for a given sequence. In mean-field theory the characteristic barrier height $\bar{F}^{\neq}$ is largely determined from the density of states and the Hamiltonian. However, in capillarity theory, the fluctuations in barrier heights are large, and $F^{\neq}$ can deviate considerably from $\bar{F}^{\neq}$.

Given a trap of energy E, the typical profile for growth of a nucleus involved in trap escape is given by Eq. **1** with $F(0)$ replaced by the energy of the trap

$$\bar{F}(N_e) = E - fN_e + \sigma N_e^z. \quad [5]$$

Given a composition with ruggedness b, the free energy $F(N)$ of the ensemble of untrapped structures is given by the total free energy $F(b, T)$ at temperature T. The value of $F(b, T)$ depends on a critical parameter,

$$b_G = T\sqrt{2s_o/c}, \quad [6]$$

which is the value of b where the system is sufficiently rugged to be glassy at biological temperature T. In terms of the total energetic variance $B^2 \equiv Ncb^2$ and the total conformational entropy $S_o \equiv Ns_o$, when $b < b_G$ the free energy $F(b, T) = -TS_o - B^2/(2T)$, and when $b > b_G$, $F(b, T)$ equals the ground-state energy $-\sqrt{2S_oB^2}$. Eq. **5** and the free energy $F(b, T)$ together determine the bulk free-energy gain f.

We approximate the surface tension as proportional to the estimate for the mean-field escape barrier per residue calculated by us earlier (16) on the correlated landscape. That calculation shows that the surface tension is always intensive and vanishes above a critical energy E^* or below a critical ruggedness b^*.

Atypical Buffed Sequences

Eq. **5** is understood to be the typical profile for trap escape for a given overall composition. However, fluctuations from this mean profile for various sequences will occur. We expect that these fluctuations can be relatively large, because reconfigurational barriers scale as $N^{1/2}$, but the distribution of a random process after N events of residues joining the growing phase has a width $\approx N^{1/2}$. Just as we found rare sequences that were anomalously rugged or funneled, we are interested now in finding rare sequences that are anomalously buffed with low kinetic barriers.

Because the variance in interaction energies is b^2, when the amount of new phase increases in a nucleation process by $\delta N_e = 1$, the change in free energy δF for a given sequence is chosen from a Gaussian distribution with mean $\delta\bar{F}(N_e)$ and variance cb^2. A particular nucleation process corresponds to a sequence of increments $\{(\partial F/\partial N_e)(N_e = 1), (\partial F/\partial N_e)(2), \ldots, (\partial F/\partial N_e)(N)\}$, such that the probability $P[F(N_e)]$ of a particular free-energy profile $F(N_e)$ is given by a path integral over N_e. The probability of a free-energy profile is analogous to the probability amplitude of finding a particular path for a quantum particle. To find the probability $\phi(F^{\neq})$ that a nucleation path has an escape barrier from a given trap lower than some value $F^{\neq}$ and that does not allow intermediate traps deeper than $F^{\neq}$ below $F(b, T)$, we introduce an infinite square-well potential $V(F(N_e)|F^{\neq})$ to constrain the profile. $V(F(N_e)|F^{\neq})$ is 0 as long as $F(b, T) - F^{\neq} < F(N_e) < E + F^{\neq}$; otherwise it is set to ∞. It serves as an absorbing boundary for any paths wandering higher than $F^{\neq}$ above E or lower than $F^{\neq}$ below $F(b, T)$. The probability we seek then can be expressed by using a Green's function,
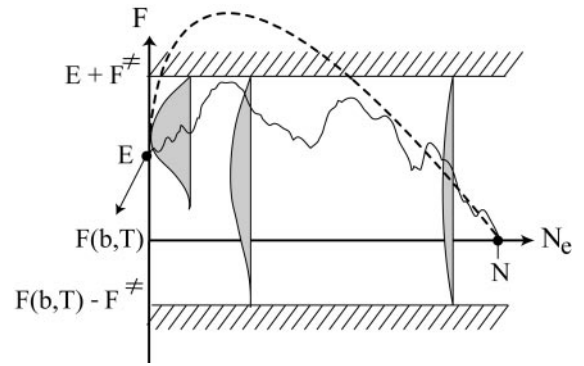
Fig. 1. Nucleation free-energy profiles in trap escape, as a function of the number of escaped residues N_e. Shown are the mean free-energy profile (dashed line) and a typical profile for sequences constrained to have barriers smaller than $E + F^{\neq} - (F(b, T) - F^{\neq})$. Also shown schematically is the propagator $G(F, N_e)$ for various values of N_e. The system rapidly relaxes to ground-state wavefunction after $O(1)$ steps (see *Appendix*).

$$G(F(b, T), N|E, 0)$$

$$= \int \mathcal{D}F(N_e)e^{-\int_0^N dN_e\left[\frac{1}{2cb^2}\left(\frac{\partial F}{\partial N_e} - \frac{\partial\bar{F}(N_e)}{\partial N_e}\right)^2 + V(F(N_e)|F^{\neq})\right]}$$

$$\cdot\delta(F(0) = E)\delta(F(N) = F(b, t)), \quad [7]$$

where the δ functions ensure that the paths must start and end at the trapped and untrapped free energies, respectively.

To calculate the appropriate fraction of nucleation paths $\phi(F^{\neq})$ that have an escape barrier from a particular trap no larger than $F^{\neq}_{MAX} = E + F^{\neq} - (F(b, T) - F^{\neq})$, we must normalize by the free propagator $G_F(F(b, T), N|E, 0)$ in the absence of absorbing walls, i.e. with V set to zero:

$$\phi(F^{\neq}) = G(F(b, T), N|E, 0)/G_F(F(b, T), N|E, 0). \quad [8]$$

Evaluating these Green's functions is made easier by using the quantum-mechanical analogy. We recognize the term $\partial\bar{F}/\partial N_e$ in Eq. **7** as a time-dependent gauge transformation in one dimension, thus we see that the propagator to free energy F after N_e steps from the path-integral problem satisfies an imaginary "time" Schrodinger equation,

$$\frac{\partial G}{\partial N_e}(F, N_e) - \frac{cb^2}{2}\frac{\partial^2 G}{\partial F^2}(F, N_e) + \dot{F}(N_e)\frac{\partial G}{\partial F}(F, N_e)$$
$$+ V(F(N_e)|F^{\neq})G(F, N_e) = \delta(F - E)\delta(N_e), \quad [9]$$

where $\dot{F} \equiv \partial\bar{F}/\partial N_e = -f + z\sigma N_e^{-(1-z)}$. The situation is shown in Fig. 1. Here $-iN_e$ plays the role of time, F plays the role of position, $\hbar = 1$, and mass $m = 1/cb^2$.

The free propagator $G_F(F(b, T), N|E, 0)$ is straightforward:

$$G_F(F(b, T), N|E, 0) = (2\pi cb^2 N)^{-1/2}e^{-(F(b,T) - \bar{F}(N))^2/2cb^2N}, \quad [10]$$

where $\bar{F}(N_e)$ is again the typical free energy profile in Eq. **5**.

We have been unable to obtain an exact, closed-form solution for the nonseparable, time-dependent problem with boundaries. However, we give in *Appendix* several analytical solutions in some reasonable limiting cases and solve the problem numerically in general.

There are polynomially many nucleation sites for a contiguous nucleus, which can be estimated geometrically. We equate this with the number of possible ways or routes N_R to escape from a trap. If any of these nucleation sites are buffed to $F^{\neq}$, the trap then is assumed to be buffed to $F^{\neq}$. We assume independent

buffing probabilities for each of the distinct routes. Then the probability that n of the routes are buffed is simply a binomial distribution for n events, the probability of each of which is $\phi(F^{\neq})$, given in Eq. **8**. Then the probability that at least one route is buffed to $F^{\neq}$, i.e. the probability that a given trap of energy E is buffed, is $p_B\,(F^{\neq}|E,\,b) = 1 - (1 - \phi(F^{\neq}))^{N_R}$.

Density of States and Density of Traps

To determine whether a sequence, rather than just a specific trap, is buffed we need to know how many candidate traps have energy E. We estimate the number of candidate traps using the mean-field calculation for a "typical" sequence. Then the effects of fluctuations sequence to sequence can be investigated.

A configurational state is a local trap if all states kinetically connected to it by a single local move are higher in energy. Because the number of states connected to a given state equals $N\nu$ when each residue can move to ν other states on average, there is a polynomial fraction of traps on uncorrelated landscapes (17), $f_T \approx 1/N\nu$. On a correlated landscape, however, there is an exponentially small fraction of traps. To see this, let the conditional probability that a state has energy E' given that it shares a fraction of q contacts with a state of energy E be $P_q(E'|E)$. Then the fraction of states at energy E that are traps is given by the expression $f_T(E) = (\int_E^\infty dE'\ P_q(E'|E))^{N\nu}$, with q suitably chosen as described below.

For a Hamiltonian consisting of interacting sets of p contacts, the conditional probability $P_q(E'|E)$ obeys a Gaussian distribution centered on $q^p E$ with variance $2B^2(1 - q^{2p})$ (18). Thus when $q \to 1$, $P_q(E'|E) \to \delta(E' - E)$, and when $q \to 0$ or $p \to \infty$, the states become uncorrelated and $P_q(E'|E) \to P(E')$. When $p = 1$, the landscape statistics of a two-body Hamiltonian are recovered. The more "many-body" the Hamiltonian is, the more decorrelated states of a given structural similarity are.

Because we are looking at states connected by single kinetic moves, we take $E' = E + \delta E$ and $q = 1 - \delta q$, with $\delta E \sim O(N)$ but small compared to E, and $\delta q \sim O(N^{-1})$, because we envision local moves of an intensive number of residues. Then the distribution of δE becomes a Gaussian with mean $pE\delta q$ and variance $4B^2 p\delta q$, and the fraction of traps becomes $f_T(E) = [(1/2)\ \mathrm{erfc}(\sqrt{p\delta q}E/2B)]^{N\nu}$ where $\mathrm{erfc}(x)$ is the complementary error function. Thus f_T is of the form of a fraction raised to a large power. It is not significant until the argument or erfc (which is intensive) is fairly large and negative. Then the error function may be asymptotically expanded around one to yield

$$f_T(E) \simeq e^{-c(E)N}, \qquad [11]$$

where $c(E) = \nu(\pi\sqrt{2}p\delta qE^2/B^2)^{-1/2}\exp(-p\delta qE^2/2\sqrt{2}B^2)$ is a function of an intensive argument. Retracing the steps in the above argument for an uncorrelated landscape (or letting $p \to \infty$) results in nearly every state constituting a trap, consistent with previous results.

The total number of traps $\Omega_T(E,\,b)$ at energy E is then the total number of states Ω times the fraction $P(E)$ of those states at energy E, times the fraction $f_T(E)$ of those states that are traps. A plot of the log number of states and log number of traps is shown in Fig. 2.

The probability that all the traps at energy E will be buffed to $F^{\neq}$ is

$$P_B(F^{\neq},\,E,\,b) = p_B(F^{\neq}|E,\,b)^{\Omega_T(E,b)}, \qquad [12]$$

assuming independent buffing probabilities for each trap. This assumption is likely to underestimate P_B. On the other hand we have assumed independent probabilities for each of the N_R routes out of a given trap to be buffed. This overestimates the value of p_B, bringing it closer to unity.
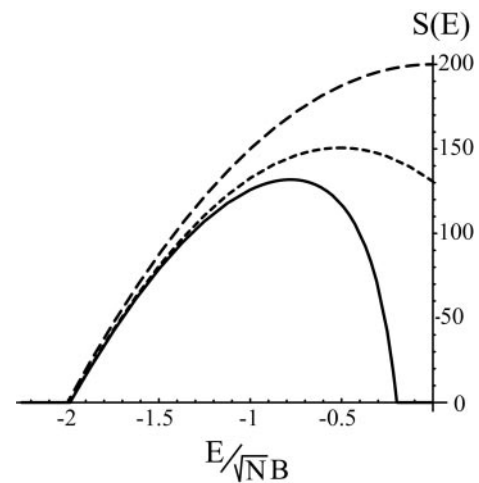

Fig. 2. Long-dashed line, log number of states vs. energy; solid line, approximate log number of traps from Eq. **11**; short-dashed line, log number of traps for the error function expression of f_T in the text above Eq. **11**. The approximation is accurate for energies near the ground state but falls off too rapidly for energies near 0. There are still traps at $E = 0$, because states may have positive energies as well. Estimates are used here for the number of states per residue ($\Omega = e^{N \times 2}$), system size ($N = 100$), and number of connected states ($\nu = 1$). Pair interactions are taken ($p = 1$), energies are in units of $\sqrt{N}B$, and $\delta q = 1/N$.

As energy increases above that of the ground state, the quantity P_B rapidly approaches a value of unity. It is hardest to buff out the lowest energies even considering that there are more traps at higher energies that must be buffed out. The fraction of sequences with available traps buffed at all energies that may be thermally occupied is

$$P_B(F^{\neq},\,b) = \prod_{E_G}^{0} P_B(F^{\neq},\,E,\,b)$$

$$= \exp\left[\int_{E_G}^{0} dE\ \Omega_T(E,\,b)\ln(p_B(F^{\neq}|E,\,b))\right]. \qquad [13]$$

Sequences with Buffed Landscapes

Let the system have nonextensive gap Δ and be rugged enough so that the ground state is thermodynamically stable at the temperature of the earth, $b > b_G$ in Eq. **6** (Fig. 3 shows a
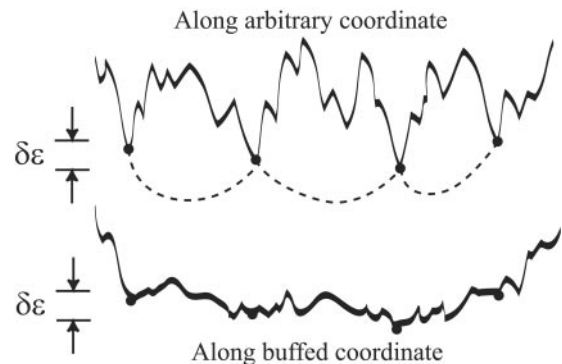

Fig. 3. Schematic of the energy landscape for a sequence with buffed ground states, projected onto a configurational coordinate. The barriers between the ground states are reduced to $F^{\neq}$, but the overall variance of energy between states is not reduced. The lowest kinetic barriers between the states determines the escape rate. Along the coordinate(s) where the landscape is buffed, the kinetic barriers between the states are reduced.

schematic of buffing between low-energy states). Note that all the thermodynamic properties of buffed landscapes are still the same as for random sequences. For example, configurational states still have variances in energies, thus there is little or no thermal entropy left in the system when $b > b_G$. It is the kinetic properties of the system that are different. Because the ground state is competing with polynomially other low-energy states, its Boltzmann weight is given by $w_N \approx (1 + c'N^\gamma e^{-\Delta/T})^{-1}$, where c' is a constant of $O(1)$, and γ is an < 1. Thus, even for a nonextensive gap $\Delta \sim O(N^{1/2})$, the system will still have large Boltzmann weight in the ground state.

How rare are these buffed sequences as system size grows larger? Using Eq. **8** and the results from *Appendix*, we find an asymptotic expression for the fraction of surviving paths. To leading order,

$$\phi \sim N^{1/2}e^{-\frac{\pi^2 cb^2}{8F^{\neq 2}}N - \frac{\sigma^2(1-z)^2}{2cb^2}N^{2z-1}}. \qquad [14]$$

The prefactor arises from the free propagator. In the exponent, the first term is a diffusion term that corresponds physically to the fact that buffed sequences are rare because of the likelihood of the specific free-energy profile to fluctuate outside of $\pm F^{\neq}$ after N steps. The second term in the exponent comes from the forcing term in the diffusion equation and embodies the fact that buffed sequences are rare because typical sequences follow the average free-energy profile. However, perhaps unexpectedly, the forcing term is subdominant, scaling more weakly with N (recall z is between $1/2$ and $2/3$, because the nucleus is roughened). Because ϕ is exponentially small, the probability that a trap is buffed over any route [in the expression for $p_B(F^{\neq}|E, b)$ above] has only polynomial corrections: $p_B \simeq N\phi$.

Because it is hardest to buff out the lowest energies, a sequence can be said to be buffed when a band of energies starting at the ground state and scaling polynomially with system size have all traps buffed to $F^{\neq}$. Buffing out polynomially many competing ground states gives (from Eq. **13**) $P_B \approx p_B e^{aN^\alpha}$ with $\alpha < 1$. This again modifies the scaling by only subdominant corrections. From Eq. **14**, more rugged sequences are harder to buff. The fraction of buffed sequences stable in their ground state at biological temperatures thus is dominated by those sequences having $b^2 = b_G^2$ in Eq. **6**. The fraction of sequences P_R having this ruggedness is given through Eq. **3**. The total fraction of buffed sequences P_{BUFF} is then given by $P_R(b_G)P_B(F^{\neq}, b_G)$ or

$$P_{BUFF} \sim e^{-\frac{(2s_oT^2 - \langle b^2 \rangle)^2}{2\sigma_b^2}N - \frac{\pi^2 s_o T^2}{4F^{\neq 2}}N}. \qquad [15]$$

For numerical values of the pair-interaction energies, such as those in the Miyazawa–Jernigan matrix, and naturally occurring amino acid abundances, sequence-to-sequence fluctuations of the barrier (second term in the exponent in Eq. **15**) dominate the possibility of finding a buffed foldable sequence, with the probability of finding a sufficiently rugged composition playing a secondary role.

From Eqs. **4** and **15** we see that the relative rareness of buffed vs. funneled sequences becomes a quantitative issue, i.e. for a gap Δ to be thermodynamically relevant on a funneled landscape, $\Delta \sim O(N)$. We address this in Fig. 4, where the fraction of sequences buffed to $F^{\neq} = 4k_BT$ is plotted as a function of chain length N as well as the fraction of sequences sufficiently gapped so that the forward folding barrier is $4k_BT$. The folding barrier is determined from a capillarity model; however, the gap still must scale extensively with chain length to ensure a constant ratio of T_F/T_G. There is a crossover in the figure beyond which it becomes easier to funnel. However, for short sequences as well as for longer sequences with degrees of freedom removed (due, for example, to native secondary or tertiary structure formation), buffing becomes more likely.
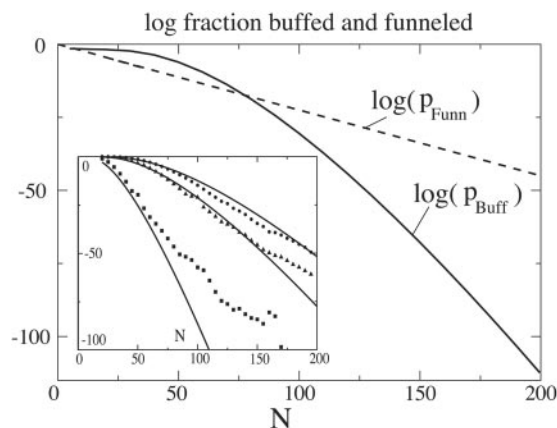
Fig. 4. The fraction of foldable sequences, on a log scale, as a function of chain length N. Both funneling and buffing mechanisms are shown here. Dashed line, fraction of funneled sequences with forward folding barrier $F^{\neq} = 4k_BT$, and $T_F/T_G = 1.6$; solid line, fraction of sequences buffed to $F^{\neq} = 4k_BT$. The adiabatic approximation for the Green's function in Eq. 9 is used (see *Appendix*). The crossover from buffing to funneling suggests the possibility of a compound mechanism for generating a functional protein sequence. The funneling mechanism removes most of the entropy while guiding the protein to a smaller ensemble of similar structures. Then for this reduced collection of states, dynamics between individual traps is most likely mediated by buffing. This dynamics may be related to functionally important motions in proteins (1). For funneled sequences we assume that no ruggedness selection occurs. We scale the Miyazawa–Jernigan interaction parameters so that in our units $T_F/T_G \approx 1.6$, a common value taken from the literature (19). Buffed sequences must be selected to be sufficiently rugged to be stable at biological temperatures and must have low kinetic barriers to be accessible on biological time scales. (*Inset*) Comparison of the fraction of buffed sequences from the adiabatic approximation and from full numerical solution to Eq. 9. Several values of ruggedness are plotted. Circle, $b = 1.255$; triangle, $b = 1.4$; square, $b = 2.0$.

Discussion

With increasing chain length it becomes exponentially rare for sequences to be significantly buffed. Likewise it is exponentially difficult to be funneled, thus the question is: Which exponential wins? Our estimates suggest that funneling overall is the dominant mechanism at least for larger proteins that fold on biological time scales. Still, the buffing mechanism raises some interesting practical questions.

Is it possible to design and synthesize a "buffed foldable protein?" Despite their rareness, buffed proteins probably exist and would be interesting objects, because they would possess a polynomial number of accessible states that could act as memories. Unfortunately, unlike funneled landscapes, the definition of buffed landscapes does not immediately lead to a design algorithm. We note that only recently has the funneling recipe been explicitly used in achieving foldability *de novo* in the laboratory (W. Jin, O. Kambara, H. Sasakawa, A. Tamura, and S. Takada, private communication). Still, combinatorial strategies are possible. Although natural protein landscapes are most likely funneled, the buffing mechanism may play a partial role in folding kinetics. Once a significant part of the protein is at least locally fixed by the funneling mechanism, the entropy is reduced to a level at which buffing can occur. This can act to remove specific intermediates or high-energy transition states on the landscape.

Other mechanisms in the spirit of buffing may exist that can facilitate folding. One can envision "antibuffed" or "gated" landscapes that have regions of phase space closed off by high-energy barriers, such that the system folds faster by exploring a smaller region of phase space, and avoiding regions that may contain deep traps. The fraction of sequences having gated landscapes may be worked out in a similar fashion as the fraction of buffed landscapes. Such a corralling mechanism on a funneled landscape is one way of

describing the gatekeeper residues discovered by Otzen and Oliveberg in protein S6 (20). Gating is similar in spirit to Levinthal's (21) original resolution to the kinetic paradoxes in protein folding as well as issues that arise when considering kinetic partitioning mechanisms (22). Gating may also play a role in avoiding misfolded protein structures responsible for aggregation-related diseases.

Funneling is a sufficient and likely but not necessary condition for achieving foldability. However, the above folding mechanisms are not mutually exclusive: Funneled proteins may also exhibit buffing. The extent to which buffing occurs in real proteins may be tested, for example, by looking at the mutational sensitivity of the weights of nonnative transient folding intermediates. Two candidates for these studies are the α-helical early-folding intermediate in β-lactalbumin (23) and the major kinetic traps in α-lactalbumin, the populations of which may be measured by disulfide scrambling (24).

The ways evolution has assisted the folding and function of proteins are manifold: We have described an alternative to funneling here. It will be interesting to see what other organizing principles of energy landscapes are important in describing structural and functional biology.

Appendix: Solutions for Green's Function for Buffing

In this section we find approximate solutions to Eq. 9 to derive analytical expressions for $\phi(F^{\neq})$ in Eq. 8. We are interested here in finding the fraction of sequences with kinetic barriers that are below $F^{\neq}$, where $F^{\neq} \sim O(k_B T)$. The wave-packet solution to Eq. 9 spreads out diffusively to width $F^{\neq}$ in a number of steps $N_e^* \sim F^{\neq 2}/b^2$. But because both $F^{\neq}$ and b are $\sim O(k_B T)$, $N_e^* \sim O(1)$. Hence we expect that by N steps the wave function will have long since relaxed to the ground state. This also means that the time rate of change of the drift term in Eq. 9 is slow compared to the rate at which the wave packet relaxes. The adiabatic parameter is given here by $|\langle 2|\partial H/\partial N_e|1\rangle/(E_2 - E_1)^2|$, where the numerator is the matrix element of $\dot{F}(N_e)$ $(\partial G/\partial F)$ between the ground and first excited eigenfunctions of the unperturbed Hamiltonian, and the denominator is the difference in the first excited and ground-state eigenvalues. A straightforward calculation gives the parameter as $L^3\sigma/\pi^4 b^4 N_e^{2-z}$ (where $L \equiv F_{MAX}^{\neq}$ is given above Eq. 8 and is of order of a few $k_B T$). This parameter is much smaller than unity over the whole range of N_e for all energies contributing to barriers to be buffed. Hence we conclude that here the adiabatic approximation is a good one. Then the time-dependent prefactor $\dot{F}(N_e)$ in Eq. 9 then may be treated as a constant parameter and the solution then given in terms of that parameter.

The adiabatic approximation is a special case of a more general transformation. We may eliminate the drift term in Eq. 9 by letting $G(F, N_e) = K(F, N_e)\exp(g(F, N_e))$, with $g(F, N_e) = \dot{F}(N_e)F/(cb^2) - (1/2cb^2)\int_0^{N_e} dN_e'\dot{F}(N_e')^2$. This yields a new differential equation for $K(F, N_e)$ with a time- and position-dependent sink term:

$$\partial_{N_e}K(F, N_e) - (cb^2/2)\partial_{FF}K(F, N_e)$$

$$+ ((\ddot{F}(N_e)/cb^2)F + V(F(N_e)|F^{\neq}))K(F, N_e)$$

$$= \delta(N_e)\delta(F - E) \lim_{N_e \to 0} e^{-\frac{\dot{F}(N_e)E}{2cb^2}}. \quad [16]$$

Because in the adiabatic approximation the coefficient of the drift term could be treated as a constant, the first term in the parentheses in Eq. 16 ($\propto \ddot{F}(N_e)$) can be neglected, because it is the time derivative of the drift coefficient. Then for $N_e > 0$, $K(F, N_e)$ satisfies a simple diffusion equation with absorbing walls such that $K(F(b, T) - F^{\neq}, N_e) = K(E + F^{\neq}, N_e) = 0$, and the initial condition is $K(F, N_e) = \delta(F - E)$. Thus the expansion in eigenfunctions may readily be written down. As we mentioned above, ground-state dominance is an excellent approximation for $N_e > 1$, thus the solution is then given by the ground-state term in the eigenfunction expansion,

$$G(F, N_e) \cong (2/L)e^{\frac{\dot{F}(N_e)}{cb^2}(F - E)}\sin(\pi(E + \delta F^{\neq})/L)$$

$$\cdot\sin(\pi(F + \delta F^{\neq})/L)e^{-\left[\frac{\pi^2cb^2}{2L^2} + \frac{\dot{F}(N_e)^2}{2cb^2}\right]N_e}, \quad [17]$$

where $\delta F^{\neq} \equiv F^{\neq} - F(b, T)$. The propagator to $(F(b, T), N)$ is then $G(F(b, T), N)$. The fraction of surviving paths is then given by $\phi(F^{\neq})$ in Eq. 8 along with Eq. 17 and the free propagator in Eq. 10.

The N scaling of the fraction of buffed sequences may be checked with an exactly solvable model. For near ground states buffed to $F^{\neq}$, the width $L = 2F^{\neq}$ does not scale with N, and thus can be replaced by a parabolic potential of a fixed, effective width. Then $V(F(N_e)|F^{\neq})$ in Eq. 16 is replaced by $\frac{1}{2}\omega^2F^2/cb^2$, where the effective frequency ω determines the spring stiffness. Here the differential Eq. 16 is equivalent to a Schrodinger equation with $\hbar = 1$, time t replaced by $-iN_e$, position replaced by free energy F, and particle mass replaced by $1/cb^2$. Shifting coordinates so the effective harmonic well is centered at 0, the solution to Eq. 16 may be obtained exactly by integrating the Lagrangian over all paths (25). The result is

$$K_{HO}(F, N_e|F_o, 0) = (\omega^{-1}2\pi cb^2\sinh(\omega N_e))^{-1/2}e^{S_{CL}(F, N_e|F_o, 0)}, \quad [18]$$

where in the quantum-mechanical analogy S_{CL} represents the (extremal) action along the classical path. Calculating this for large N shows that, analogous to the "forcing term" we saw before, the classical action $S_{CL}(F, N|F_o, 0)$ does not dominate the scaling behavior. Instead the prefactor, measuring deviations from the path of least action, plays the most important role, scaling as $\sim e^{-\omega N/2}$ for large N. Following the same arguments as before, this is essentially the scaling law for the fraction of buffed sequences. Comparing the scaling law here with that in Eq. 15, we see that the effective frequency ω is such that the unforced propagators in the hard-wall and parabolic-wall cases have the same survival probability after N steps.

1. Frauenfelder, H., Sligar, S. G. & Wolynes, P. G. (1991) *Science* **254,** 1598–1603.
2. Leopold, P. E., Montal, M. & Onuchic, J. N. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 8721–8725.
3. Bryngelson, J. D. & Wolynes, P. G. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 7524–7528.
4. Ueda, Y., Taketomi, H. & Gō, N. (1975) *Int. J. Peptide Protein Res.* **7,** 445–459.
5. Onuchic, J. N., Socci, N. D., Luthey-Schulten, Z. & Wolynes, P. G. (1996) *Fold. Des.* **1,** 441–450.
6. Clementi, C., Jennings, P. A. & Onuchic, J. N. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 5871–5876.
7. Takada, S. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 11698–11700.
8. Govindarajan, S. & Goldstein, R. A. (1997) *Proteins Struct. Funct. Genet.* **29,** 461–466.
9. Villain, J. (1985) *J. Phys.* **46,** 1843–1852.
10. Kirkpatrick, T. R., Thirumalai, D. & Wolynes, P. G. (1989) *Phys. Rev. A* **40,** 1045–1054.
11. Pande, V. S., Grosberg, A. Y. & Tanaka, T. (1997) *Biophys. J.* **73,** 3192–3210.
12. Miyazawa, S. & Jernigan, R. L. (1996) *J. Mol. Biol.* **256,** 623–644.
13. Derrida, B. (1981) *Phys. Rev. B* **24,** 2613–2626.
14. Šali, A., Shakhnovich, E. & Karplus, M. (1994) *J. Mol. Biol.* **235,** 1614–1636.
15. Wolynes, P. G. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 6170–6175.
16. Wang, J., Plotkin, S. S. & Wolynes, P. G. (1997) *J. Phys. I [French]* **7,** 395–421.
17. Bryngelson, J. D. & Wolynes, P. G. (1989) *J. Phys. Chem.* **93,** 6902–6915.
18. Plotkin, S. S. & Onuchic, J. N. (2002) *Q. Rev. Biophys.* **35,** 205–286.
19. Onuchic, J. N., Wolynes, P. G., Luthey-Schulten, Z. & Socci, N. D. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 3626–3630.
20. Otzen, D. E. & Oliveberg, M. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 11746–11751.
21. Levinthal, C. (1969) in *Mossbauer Spectroscopy in Biological Systems*, eds. DeBrunner, P., Tsibris, J. & Munck, E. (Univ. of Illinois Press, Urbana), pp. 22–24.
22. Thirumalai, D. (1995) *J. Phys. I [French]* **5,** 1457–1467.
23. Kuwata, K., Shastry, R., Cheng, H., Hoshino, M., Batt, C. A., Goto, Y. & Roder, H. (2001) *Nat. Struct. Biol.* **8,** 151–155.
24. Chang, J.-Y. (2002) *J. Biol. Chem.* **277,** 120–126.
25. Feynman, R. P. & Hibbs, A. R. (1965) *Quantum Mechanics and Path Integrals* (McGraw–Hill, New York).