Speeding Protein Folding Beyond the Gō Model: How a Little Frustration Sometimes Helps

Steven S. Plotkin*
Department of Physics, University of California, San Diego, California

ABSTRACT **By perturbing a Gō model toward a realistic protein Hamiltonian by adding non-native interactions, we find that the folding rate is in general enhanced as ruggedness is initially increased, as long as the protein is sufficiently large and flexible. Eventually, the rate drops rapidly toward zero when ruggedness significantly slows conformational transitions. Energy landscape arguments for thermodynamics and kinetics are coupled with a treatment of non-native collapse to elucidate this effect. Proteins 2001;45:337–345.**

Key words: Gō model; folding rate; ruggedness; non-native collapse

INTRODUCTION

Theorists seek to capture the essence of protein folding with simple models of a self-interacting polymer chain.[2,8,9,14,15,29,31,39,47,49] There are two distinct limits pertaining to the nature of the interactions in this minimalist approach. One is that of purely random interactions and is considered too frustrated to describe real proteins. Another is the Gō model,[46] for which the polymer is self-attractive only for those interactions present in the native configuration. This is considered too unfrustrated to describe real proteins and also impossible to achieve in practice. Because these two models bracket the behavior of real proteins, we consider perturbing from the Gō model toward real protein interactions by adding some non-native noise.

Gō models have been implemented to describe the folding mechanism of real proteins with considerable success,[1,5,6,12,18,26,35,38,41,42] indicating that many real proteins have interactions that are minimally frustrated enough so that a Gō model can be an accurate description.

Some of the effects of adding non-native interactions on the folding mechanism for the Honeycutt-Thirumalai β-barrel model[19] were investigated in Refs. 28 and 40, and specific non-native interactions in the transition state were examined for a lattice model in Ref. 23. Local helical propensity was enhanced for α-spectrin SH3, a β-sheet protein, in Ref. 37.

In this article we investigate the general effects on folding rate when non-native interactions are superimposed on a Gō potential.

At first glance, one would expect that adding frustration begins to slow the rate at the transition temperature or, at best, has initially no effect. What follows is a derivation of the somewhat counterintuitive result that in general the folding rate initially increases as ruggedness is increased from zero. Eventually of course, the rate decreases drastically, so a plot of the folding rate versus non-native interaction strength should look like Figure 1. Then the question of where real protein interactions reside on this plot may be addressed. For some fast-folding proteins, it is possible that non-native noise in the system may actually assist folding.

THERMODYNAMICS

Consider first the thermodynamics of a protein obtained from a statistical analysis of a correlated landscape.[33,34,36] The energy, entropy, and free energy as functions of the fraction of native contacts Q, are given by†

$$E(Q) = QE_N - \frac{\Delta^2(Q)}{T}(1 - Q) \tag{1a}$$

$$S(Q) = S_c(Q) - \frac{\Delta^2(Q)}{2T^2}(1 - Q) \tag{1b}$$

$$F(Q) = QE_N - TS_c(Q) - \frac{\Delta^2(Q)}{2T}(1 - Q). \tag{1c}$$

These quantities are shown in Figure 2, and the parameters used in them are given in Table I. $S_c(Q)$ in Eq. (1b) is the configurational entropy in the system versus Q, E_N is the extra internal energy in the native state (the stability gap), and $\Delta^2(Q)(1 - Q)$ is the non-native variance, which is a measure of the overall ruggedness of the energy landscape [see Eq. (3)]. The variance $\Delta^2(Q)(1 - Q) \to 0$ as $Q \to 1$. In obtaining the functional form of the non-native ruggedness, it is assumed here that all the native contacts have roughly the same strength.**

The energy in the native state is the number of native contacts M times the mean native attraction energy ϵ ($\epsilon < 0$). If N is the number of interacting residues in the

Grant sponsor: National Science Foundation Bio-Informatics Fellowship; Grant numbers: DBI9974199 and 0084797.

*Correspondence to: Steven S. Plotkin, Department of Physics and Astronomy, University of British Columbia, 6224 Agricultural Road, B.C. V6T 1Z1, Canada. E-mail: steve@physics.ubc.ca

Received 21 August 2000; Accepted 10 July 2001

†We generally set Boltzmann's constant $k_B = 1$ in this article, so temperatures have units of energy, and entropies are in units of k_B.

**When there is variance in the native energies, the non-native ruggedness terms are proportional to $(\Delta^2(Q) + Q\Delta_N^2(Q))(1 - Q)$, where Δ_N^2 is the native variance.[34] If the set of native energies has variance but the distribution is fully specified, then the ruggedness terms are again proportional to $\Delta^2(Q)(1 - Q)$.[35]

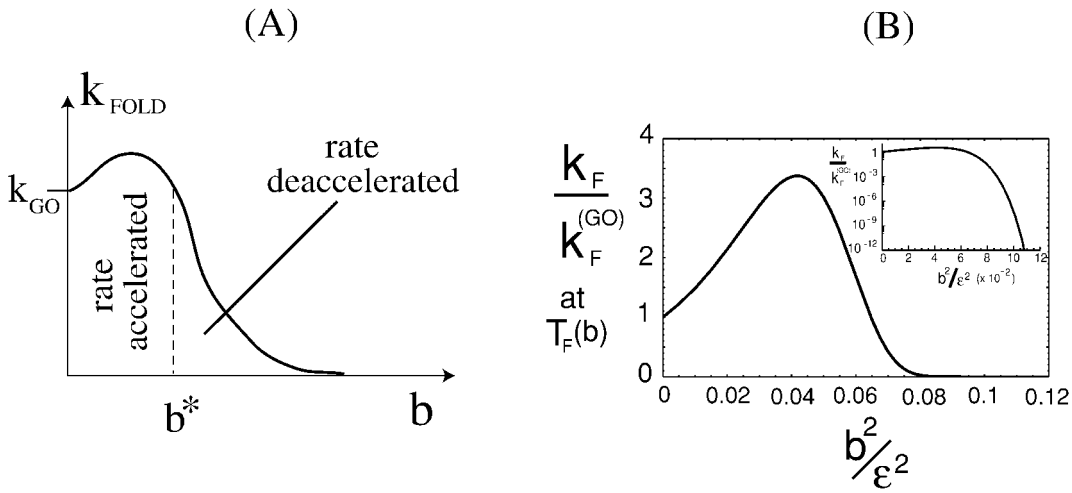

Fig. 1. Rate versus non-native interaction strength is split up into two regimens: one where it assists folding, and the other where it hinders. **A:** Schematic depicting the two regimens. **B:** Result of the theoretical model introduced in the text, for a system with parameters given in Table I. The rate plotted here is the folding rate at T_F, which is itself a function of b [see Eq. (9)], that is, stability is fixed here, temperature is not. However T_F changes by $\lesssim 5\%$ over the range of this plot. Initially, the rate rises as $\sim Nb^2/\epsilon^2$, then strongly decreases for larger b as non-native interactions slow conformational transitions. The inset of (B) shows a semilog plot of the same rate; it can be seen that there is a regime where the rate is roughly constant as ruggedness is increased from zero and then a turnover where the rate drastically decreases.

polymer chain and z is the mean number of contact interactions per residue,

$$E_N = M\epsilon = zN\epsilon. \tag{2}$$

The overall non-native ruggedness $\Delta^2(Q)(1 - Q)$ is given by

$$\Delta^2(Q) = Mb^2\eta(Q), \tag{3}$$

where $\eta(Q)$ is the non-native packing density ($0 < \eta(Q) < 1$), b^2 is the intrinsic variance per non-native interaction, and M is the total possible number of (non-native) interactions, that is, the native state is assumed to be fully collapsed with the maximal number of contacts, and this is the maximal number of total contacts of any state. The density $\eta(Q)$ tends to increase upon folding (see The Collapse Transition), hence the ruggedness scale $|\Delta(Q)|$ increases as well. The strength b of non-native interactions is taken to be weak:

$$\frac{b}{\epsilon} \ll 1, \tag{4}$$

therefore, the ratio of folding transition temperature T_F to thermodynamic glass temperature T_G is large

$$\frac{T_F}{T_G} \gg 1, \tag{5}$$

that is, the proteins we consider are strongly (but not infinitely) unfrustrated—we are perturbing away from the Gō model.

The configurational entropy $S_c(Q)$ has the property that entropy loss on folding is more rapid initially than in later stages. In mean-field theories, this arises from the entropy loss to close larger loops initially, and in capillarity theories this arises from the additional surface entropy cost to form a given native nucleus. We approximate this nonlinear effect here by assuming an entropy of the form

$$S_c(Q) = S_o(1 - Q) - \mathrm{Tent}(Q), \tag{6}$$

where $S_o \equiv Ns_o$ is the total conformational entropy in the unfolded ($Q = 0$) state (s_o is the log number of conformational states per residue), and Tent(Q) is a tent function:

$$\mathrm{Tent}(Q) = \begin{cases} 2\phi Q & Q < 1/2 \\ 2\phi(1 - Q) & Q > 1/2 \end{cases}. \tag{7}$$

We have let the barrier be at $Q^{\neq} = 1/2$ for simplicity of argument.

At the transition temperature T_F, the free energy of the folded and unfolded states are equal:

$$F(0) \cong F(1)$$

$$-T_F S_o - \frac{\Delta^2(0)}{2T_F} \cong E_N. \tag{8}$$

The folding temperature is a decreasing function of b. From Eqs. (3), (4), and (8):

$$T_F(b) \approx T_F(0)\left(1 - \frac{s_o\eta(0)}{2z}\frac{b^2}{\epsilon^2}\right), \tag{9}$$

where $T_F(0) = -z\epsilon/s_o$. At fixed temperature, folding becomes more uphill as non-native interactions are added, and this tends to slow down the rate. However, the quantity we are investigating here is the rate at folding equilibrium, that is, where the native and unfolded states have equal stability. The transition temperature that induces this scenario is a function of b; however, T_F varies by less than a few percent over the whole range of Figure 1.

Using Eq. (8) in Eq. (1c) gives

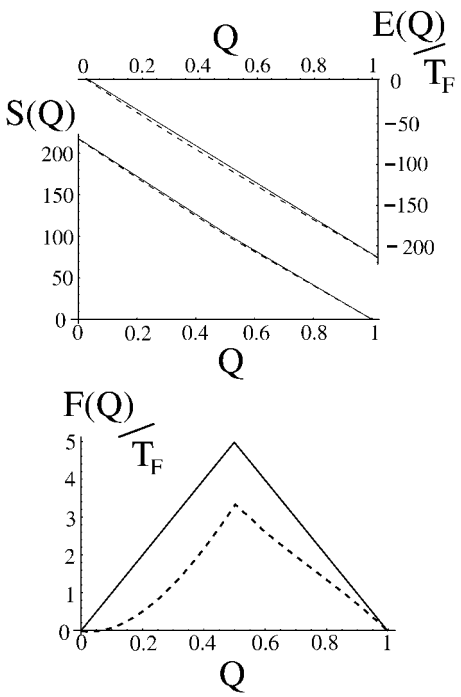

Fig. 2. Energy, entropy, and free energy versus Q in the model, given by Eq. 1. Parameters used in the model are given in Table I. For the Gō model $\Delta^2 = 0$ (solid line). Then the energy $E(Q) = QE_N$ is a linear function of Q, and the entropy $S(Q)$ is given by Eqs. (6) and (7) with $\phi = 5$, and so is a bilinear function of Q. However, the bilinearity is weak: $\phi/S_o \approx 0.02$ and is nearly indiscernible here; the free-energy barrier arises from the incomplete cancellation of large terms. Because of the bilinear approximation for the configurational entropy and pairwise additive energy function, the Gō free energy has a tent functional form at the folding transition temperature T_F. When ruggedness is introduced, when $\Delta^2 > 0$ (dashed line). We took $b^2 \cong 0.04\,\epsilon^2$ [see Eq. (3); this is where the folding rate in Fig. 1 is maximal). The non-native density function $\eta(Q)$ used here is a fit to the off-lattice data in Figure 4(C). When $\Delta^2 > 0$, the energy is lowered by $\delta E = -Mb^2\eta(1 - Q)/T$ from Eq. (1a). From Eq. (1b), the entropy is lowered by $T\delta S = -\delta E/2$, so the free energy is everywhere lowered by $\delta F = \delta E - T\delta S = -\delta E(Q)/2$. The folding free-energy barrier at T_F is *lowered* by non-native ruggedness, because for the non-native density $\eta(Q)$ is greater in the transition state than the unfolded state.

$$\left.\frac{F(Q) - F(0)}{T}\right|_{T_F} = \mathrm{Tent}(Q) - \frac{Mb^2(1 - Q)}{2T_F^2}\,[\eta(Q) - \eta(0)]. \tag{10}$$

When $b = 0$ the free energy at T_F is the tent function (see Fig. 2), and so ϕ in Eq. (7) is thus $\Delta F^{\neq}(b = 0)/T_F$. Then from Eq. (10) the free energy barrier at T_F is given by

$$\frac{\Delta F^{\neq}(T_F)}{T_F} \approx \frac{\Delta F^{\neq}(b = 0)}{T_F} - \frac{Mb^2}{4T_F^2}\,\Delta\eta^{\neq}, \tag{11}$$

where $\Delta\eta^{\neq} = \eta(Q^{\neq} = 1/2) - \eta(0)$ is the change in non-native density between the barrier peak and unfolded state ($\Delta\eta^{\neq} < 1$). So long as $\Delta\eta^{\neq} > 0$, the barrier height decreases as non-native variance (b^2) increases, as shown in Figure 2. We now show that this is generically the case by considering the physics of collapse for our problem in question.

The Collapse Transition

In this section we investigate the coupling of non-native density with the amount of native structure present in a protein and show that native topological constraints can induce a collapse transition on the non-native parts of the protein. Then the trend in Eq. (11) of adding non-native ruggedness would be to lower the folding barrier.

Collapse occurs below a temperature T_θ, defined as the temperature where the free energy of the coil and collapsed molten globule phases (both at $Q \cong 0$) are equal:

$$F_{\mathrm{coil}}(T_\theta) = F_{\mathrm{mg}}(T_\theta). \tag{12}$$

Again using Eq. (1c), but now noting that the conformational entropies are different in the coil and globule phases, and that $\eta \cong 1$ in the globule and $\eta \cong 0$ in the coil phase, we have

$$-T_\theta S_{\mathrm{coil}} = -T_\theta S_{\mathrm{mg}} - \frac{Mb^2}{2T_\theta} - Ma. \tag{13}$$

Note we have now allowed for a mean homopolymer attraction a in general, for reasons that will become clear below. Using the reduction in entropy for collapsed versus coil chain statistics[4,27]

$$S_{\mathrm{coil}} - S_{\mathrm{mg}} = N \log \nu - N \log(\nu/e) = N \tag{14}$$

gives for the collapse temperature

$$T_\theta = \frac{za}{2}\left(1 + \sqrt{1 + \frac{2b^2}{za^2}}\right). \tag{15}$$

Note from Eq. (15) that when $b = 0$, $T_\theta = za$, that is, the collapse temperature is the mean homopolymer attraction times the number of contacts per residue. When $a = 0$, $T_\theta = b\sqrt{z/2}$, that is, non-native heterogeneity can drive collapse, with the collapse temperature now scaling with the root number of contacts per residue times the width of interactions.

Now we note that in our model (Gō perturbed by weak non-native heterogeneity, with $a = 0$), collapse and folding will tend to occur together, with folding driving the collapse through native structural constraints. So the *total* density increases from zero to one as Q increases from 0 to 1. Moreover, the non-native density $\eta(Q)$ should increase as well, because non-native polymer is more strongly constrained by larger native cores (see Fig. 3). The simplest approximation to capture this increase in density upon folding is to replace the mean homopolymer field a by the native energy scale ϵ times the fraction of native bonds made Q:

$$a(Q) = \epsilon Q. \tag{16}$$

This is the effective homopolymer field for the ensemble of states with fraction Q of native structure. Using Eq. (16) in Eq. (15) and noting that the glass temperature

$$T_G = \sqrt{\frac{zb^2}{2s_{\mathrm{mg}}}} \tag{17}$$

gives

$$T_\theta = T_F\left(\frac{s_o Q}{2} + \sqrt{\left(\frac{s_o Q}{2}\right)^2 + s_{\mathrm{mg}}\left(\frac{T_G}{T_F}\right)^2}\right), \tag{18}$$

TABLE I. Parameters in the Model

Model	Polymer length (N)	Contacts per residue (Eq. 2) (z)	Conformational entropy per residue (Eq. 6) (s_o)	Entropy nonlinearity (Eq. 7) (ϕ)	Native contact energy (Eq. 2) (ϵ)	RMS non-native contact energy (Eq. 3) (b)	Folding transition temperature (Eq. 8) (T_F)
Gō, Gō + non-native	64	1.25	3.4	5.0	−1.0	0, b	0.37

where s_o and s_{mg} are the entropy per residue in the coil and globule state, respectively. Note that in Eq. (18), $T_\theta > T_F$, that is, the non-native parts of the protein are in a collapsed state at folding equilibrium, as long as the term in parentheses is greater than one, and T_F is below the collapse temperature. This gives a critical value Q_θ where collapse occurs during folding, that is, when $Q \gtrsim Q_\theta$, $\eta \approx 1$ and when $Q \lesssim Q_\theta$, $\eta \approx 0$. This is sketched in Figure 4(A) below. Solving for Q_θ gives

$$Q_\theta = \frac{1}{s_o} - \frac{s_{mg}}{s_o}\left(\frac{T_G}{T_F}\right)^2 = \frac{1}{s_o} - \frac{s_o b^2}{2z\epsilon^2} \cong \frac{1}{s_o}, \tag{19}$$

and Eq. (18) can be rewritten as

$$T_\theta(Q) \cong Q s_o T_F = Q z \epsilon. \tag{20}$$

The approximations in Eqs. (19) and (20) are valid by construction of the problem, that is, because Eq. (4) holds. For example, if b/ϵ is about 1/20, then the second term in Eq. (19) is of order 1/400 and can be neglected. Equation (19) says that the more chain entropy the polymer has (the more flexible it is), the *sooner* it collapses when folding at the transition temperature.

Calorimetric measurements of the conformational entropy change per residue in unfolding to the coil state for say Barnase give $s_o \cong 55$ J/K · mol residue $\cong 6.8\, k_B$ per residue.[25] This entropy also counts side chain conformational entropy, which is estimated to be about 13 J/K · mol residue $\cong 1.6\, k_B$ per residue,[10] giving a net chain conformational entropy of about $5.2\, k_B$ per residue in the coil state; therefore, $Q_\theta \cong 0.2$. For typical off-lattice simulations,[5,28] $s_o \cong 3.4\, k_B$; therefore, $Q_\theta \cong 0.3$. So collapse typically occurs before the barrier is reached [see Fig. 4(C)]. For lattice simulations $Q_\theta \cong 0.6$, which is around $Q^{\neq}$. In any event, $\eta(Q^{\neq})$ will tend to be greater than $\eta(0)$ as long as the system is large enough and bulk thermodynamics can be used (see Fig. 4), (however, see also caveats in Appendix A). The values of non-native packing density obtained from simulations are smaller than 1, probably because of finite size and stiffness effects. Applying bulk thermodynamics to the residual segments of non-native polymer may not be an accurate approximation in some cases.

More complete treatments of the coupling of density with native similarity can be made within the energy landscape framework.[33] It is fairly straightforward to write an approximate free energy as a function of both η and Q and then minimize with respect to η to obtain the density as a function of Q. We have taken the simplest approach here to illustrate the coupling of collapse with the thermodynamics. Some cautionary notes are made in Appendix A regarding a possible reversal of the trend on barrier height in small, stiff proteins, or proteins with a significant amount of homopolymeric self-attraction.

Finite-size and stiffness effects are playing a role in Figures 4(B) and 4(C): the density never reaches unity. The degree of coarse graining present in the lattice model is such that non-native polymer may not be treated as Gaussian loops, and $\eta(Q)$ in Fig. 4(B) does not show an increase with Q. However, in the larger and more flexible off-lattice models, the increase in non-native density with folding is clearly present, and $\Delta\eta^{\neq} > 0$. Because the increase is less than 1, we keep $\Delta\eta^{\neq}$ as a parameter in the equations below for folding rates.

The point of the above arguments is to show that large, flexible polymers show at least partial collapse before the transition state, and by Eq. (11) this leads to a reduction in free energy barrier as non-native interaction strength is increased from zero.

Kinetics

What about the folding rate? The question is now whether the slowing of the prefactor is a larger effect than the decrease in barrier height, as we add non-native ruggedness. Because the ruggedness is weak, the kinetics are single exponential (there is a single dominant folding barrier), and a Kramers law holds for the rate:

$$k \cong \tau^{-1}(b) e^{-\Delta F^{\neq}(b)/T} \tag{21}$$

where the prefactor is proportional to the reciprocal of the reconfiguration timescale.[3,34,44,48]

If we were to follow the argument for the dependence of the prefactor on ruggedness for an uncorrelated landscape,[3] or for a correlated landscape at low temperature with activated dynamics,[48] we would find that the ratio of rates is given by

$$\left.\frac{k(b)}{k(b=0)}\right|_{T_F} = \exp\left(\frac{Mb^2}{4T_F^2}\Delta\eta^{\neq} - \frac{Mb^2}{T_F^2}\mathcal{G}\right). \tag{22}$$

The first term in the exponential comes from Eq. (11), the second from the theory for reconfiguration time $\tau(b)$. The parameter $\mathcal{G}$ is a positive function of T/T_G on the uncorrelated landscape, and on the correlated landscape is a function of both T/T_G and structural entropic factors having to do with the density of states of given similarity to a trap. So by inspection of Eq. (22), in this low temperature regime, the rate may go up or may go down with non-native interaction strength, depending on the relative strength of the two terms in the exponent.

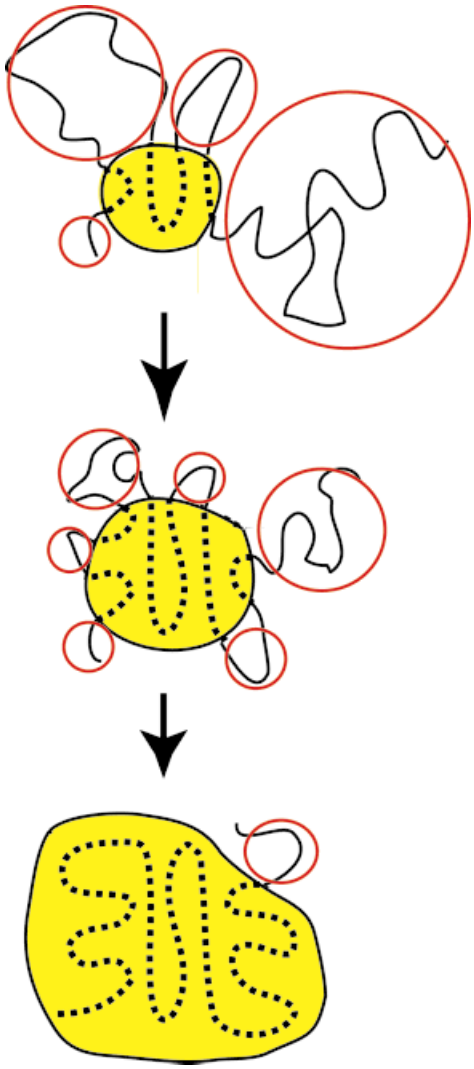

Fig. 3. As folding progresses, the non-native polymer halo surrounding the native core (central yellow globule) has more topological constraints placed upon it. Therefore, the non-native packing density, given by the total number of non-native contacts divided by the characteristic volume of non-native polymer (red spheres), tends to increase with folding. This increase in density is responsible for the stronger effect of non-native interactions on the free energy as Q increases.

However, an important result arising from energetic correlations in the landscape is the existence of a critical temperature T_A where the dynamics becomes unactivated.[48] Above this temperature, the dynamics is similar to reconfigurations in a normal liquid rather than the hopping dynamics of trap escape in a supercooled liquid or glassy system,[17,21,45,48] that is, above T_A, the prefactor $\tau^{-1}(b)$ remains nearly constant with ruggedness, because at these temperatures, the Rouse modes depend much more weakly on the ruggedness introduced. The existence of such a temperature scale can be seen from the following simple argument.[48] We can think of escape from a local trap as a mini-unfolding event: escape is driven by entropy and is opposed by the putative trap's low energy, say E_i. Then, as in unfolding, the escape barrier arises from a mismatch between entropy gains and energy losses as the system reconfigures out of the trap, so we can rewrite Eq. (1c) for the free energy relative to the state i as

$$F(q) = qE_i - TS_c(q) - \frac{\Delta^2}{2T}(1-q), \tag{23}$$

where E_N in Eq. (1c) is replaced by E_i, Q in Eq. (1c) is replaced by q, defined as the fraction of contacts shared with the trap state i, and density changes during untrapping are not particularly important since $T_A < T_\theta$ (energetic trapping occurs only when the polymer is collapsed.[4]) The transition to unactivated dynamics occurs when the states typically occupied at that temperature have zero escape barrier. Setting E_i in Eq. (23) to the thermal energy of states at temperature T,*

$$E_i \approx \bar{E}(T) = -\frac{\Delta^2}{T}, \tag{24}$$

we note that T_A occurs when the free energy profile is downhill away from the trap at $q = 1$, that is, when $\partial F_> / \partial Q = 0$ in our model, where the subscript $>$ indicates the high q region of the piecewise free energy function, eq. (23) [Eq. (6) for the configurational entropy has a piecewise structure]. Using Eqs. (7), (23), and (24), this gives

$$T_A = T_G\left(1 - \frac{2\phi}{S_o}\right)^{-1/2} \tag{25}$$

for the transition temperature to activated dynamics. From Eq. (25) we see that $T_A > T_G$ by an amount that depends on the deviation from linear entropy loss over the total unconstrained entropy, that is, by the entropic contribution to the barrier. There is no energetic contribution because we have used q as the order parameter and assumed pairwise interactions.†

For model proteins of size $N \sim 100$, $T_A \approx (1.2 - 1.8)T_G$. A plot of the escape time on a correlated landscape is given in Figure 5. Because of the large entropy and relatively small barrier in our model $T_A \approx 1.02\, T_G$ here. But in any event, for the regimen we are interested in, $T_G/T_F \lll 1$ [c.f. Eq. (5)], it is also true that

$$\frac{T_A}{T_F} \lll 1, \tag{26}$$

and so the characteristic temperatures where folding occurs ($\sim T_F$) are way above the transition temperature for activated diffusion by construction of the problem; see Figure 5.

Expanding Eq. (21) around $b = 0$ then amounts to setting $\mathcal{G} = 0$ in Eq. (22). That is, we can neglect the

*C.f. Eq. (1a) at $Q = 0$. For strata of states with $Q > 0$, the larger ruggedness scale $\Delta(Q)$ increases T_A and T_G for that stratum of states.

†One must be consistent in interpreting Eq. (25). In mean-field theory, ϕ is extensive and $T_A > T_G$ in the thermodynamic limit. But in the capillarity theory, the entropic deviation ϕ comes from surface entropy and should scale as $N^{2/3}$ (or even with a smaller power if the interface is roughened). One might argue that since $S_o \sim N$, T_A approaches T_G as $N \to \infty$, but matching the theories in this fashion is incorrect because Eq. (23) is not valid in the capillarity limit. In the capillarity theory, a dynamical transition can only be seen by investigating where the intensive surface tension vanishes.

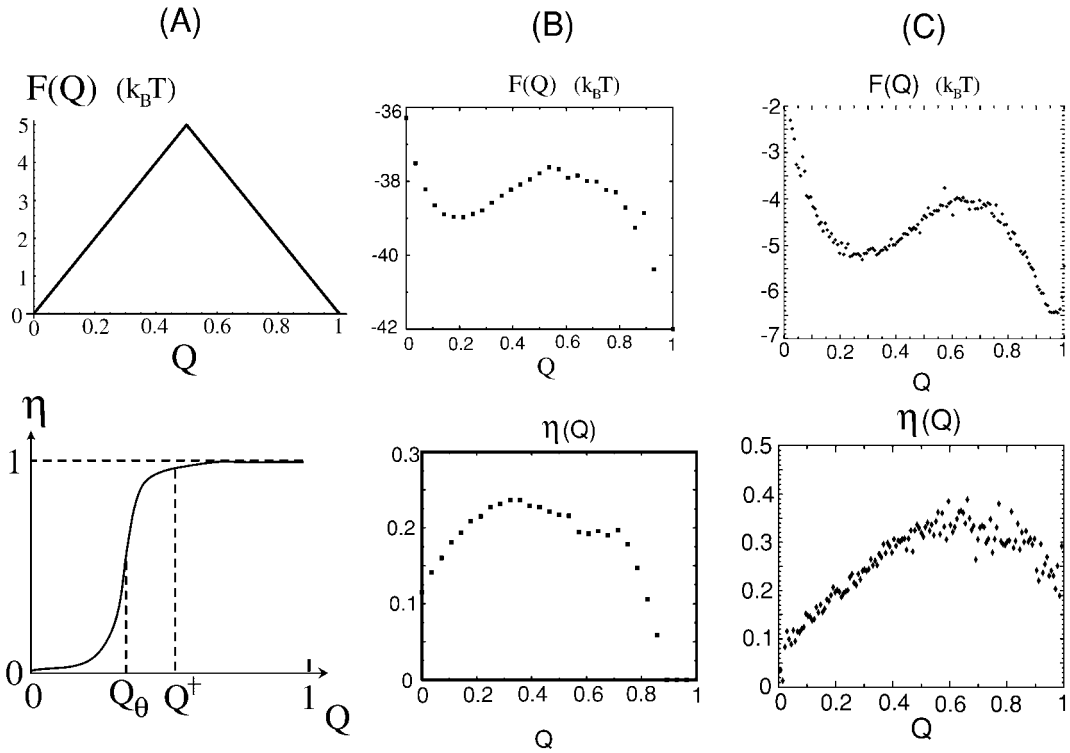

Fig. 4. Free-energy profiles $F(Q)$ and non-native polymer density $\eta(Q)$ in the model for (**a**) the simple bulk-mean-field model, (**b**) a 27-mer lattice model, and (**c**) an off-lattice model for the 57-residue fragment corresponding to the α-spectrin *SH3* domain (PDB code 1BK2). Simulation data are taken from Gō models. (a) Collapse occurs at Q_θ before the barrier peak at $Q^{\neq}$. The transition is significantly rounded for typical sized proteins, as in (b) and (c), and so the important quantity to investigate is the density change $\Delta\eta^{\neq}$ between the unfolded and transition state. The quantity $\Delta\eta^{\neq}$ is thus retained explicitly in the equations for barrier heights and rates in the main text. In (b), the non-native density is overall fairly small and is comparable in the unfolded and transition state, $\eta(Q_U) \approx \eta(Q^{\neq}) \approx 0.22$. Thus, Eq. (10) gives a barrier height roughly independent of b at least for small b. In (c), on the other hand, the non-native density rises to values larger in overall magnitude and is monotonically increasing until the barrier peak: $\eta(Q^{\neq}) \approx 0.35$ and $\eta(Q_U) \approx 0.2$. For the parameter values in Table I, Eq. (10) then gives a barrier height decreasing with b as $\Delta F^{\neq}/T_F \approx -10b^2/\epsilon^2$. The drop-off at high Q in (B) and (C) is most probably due to stiffness effects on the small pieces of residual non-native polymer in this regime.

slowdown in prefactor as ruggedness increases, to the first approximation, due to the existence of a dynamic glass temperature (see Fig. 5).

Inserting Eq. (9) into Eq. (22) and expanding for small b gives the forward rate at the folding transition temperature for weak non-native interactions:

$$\left.\frac{k}{k_{GO}}\right|_{T_F} \cong 1 + M\Delta\eta^{\neq}\frac{s_o^2}{4z^2}\left(\frac{b^2}{\epsilon^2}\right), \tag{27}$$

where $\Delta\eta^{\neq} = \eta(Q^{\neq}) - \eta(0)$ and we have used $Q^{\neq} \cong 1/2$. Equation (27) is the central result of this article, that is that the forward folding rate shows an initial increase as random non-native interactions are added to the Hamiltonian. The effect on rate due to the change of T_F with ruggedness b is a higher order effect ($\mathcal{O}(b/\epsilon)^4$) than the dependence of the barrier in Eq. (27).

The increase in rate occurs until around $b_A - \delta b(N)$, where b_A is where $T_F(b) \cong T_A(b)$ (about 0.3 here), and $\delta b(N)$ is the finite-size fluctuation of b_A due to temperature fluctuations which round the transition.[22] That is,

$$\delta b \approx \delta T_A = \left.\frac{T}{C_v}\right|_{T_A} = \left.\frac{T^2}{\sqrt{M}b^2}\right|_{T_A} \approx \frac{b_A}{\sqrt{N}}, \tag{28}$$

where Eq. (1b) was used for the entropy at $Q = 0$. This gives a value for $\delta b \approx 0.02 - 0.04$.

Realistic values of b^2/ϵ^2 for a typical protein can be obtained from the ratio of folding to glass temperature, given by

$$T_F/T_G = \lambda + \sqrt{\lambda^2 - 1}, \tag{29}$$

where $\lambda = \sqrt{z/2s_o}(\epsilon/b)$.[16] Commonly accepted values of T_F/T_G for proteins are about $1.6 - 2.0$.[30] Using the values in Table I for s_o and z gives $(b^2/\epsilon^2) \approx (0.1\text{–}0.15)$, which is above the rate enhancement regimen (see Fig. 1). If the effects of non-native rate enhancement are observable, they will be seen possibly in only the faster folding proteins. On the other hand, it was recently argued on calorimetric grounds that T_F/T_G may exceed 6 for some two-state proteins,[20] and for such proteins rate enhancement would be observable, in principle by combining calorimetric and rate measurements. Such an observation

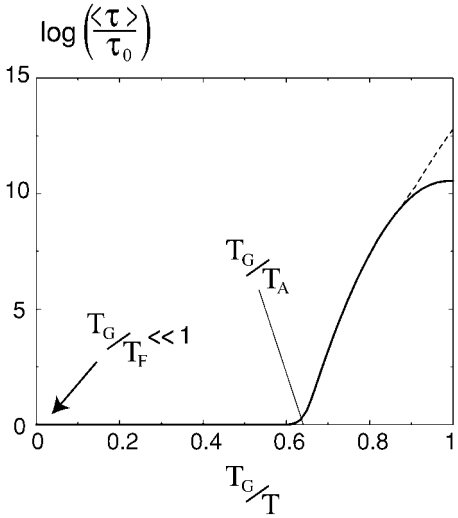

Fig. 5. Log of the reconfiguration time versus reciprocal temperature in units of T_G, for a system of size $N = 64$, adapted from Ref. 48. This is used in Eq. (21) to produce the rate curve in Figure 1(B). At temperatures above T_A, the time to reconfigure is $\sim\tau_o$, below T_A the time increases exponentially as $\exp(f(T_A/T)N)$ with $f(x) = 0$ for $x < 1$. The width of the transition $\to 0$ as $N \to \infty$ and the value of $T_A \to \approx 1.8$ as $N \to \infty$ for the mean field correlated landscape, with the numerical value of the entropy taken from simulations. When $T \lesssim T_G$ for a finite system, the deepest obligatory trap tends to dominate the kinetics, and the relaxation rate turns over to an Arrhenius law with slope corresponding to the barrier height for escape from that trap (shown schematically by the dashed line).

would support the existence of a dynamic glass transition in protein systems.

CONCLUSIONS

As the variance in non-native interactions is increased from zero, the folding rate initially increases by a factor of about 2–4 and then eventually drops rapidly toward zero (see Fig. 1). There is a regime near the Gō Hamiltonian where, on a log scale, the rate is relatively robust to changes in non-native interaction strength. The density of non-native polymer must be greater at the barrier peak than in the unfolded state for rate enhancement to be observed. Moreover, rate enhancement implies a regime where dynamics are relatively unaffected by changes in ruggedness due principally to the existence of a dynamic glass temperature.

Why does the rate initially increase? The upshot is as follows. First, it follows from energy landscape theory that if there is no change in density upon folding, then changes in ruggedness do not affect the barrier height for a well-designed protein, as can be seen from Eq. (11): if $\Delta\eta^{\neq} = 0$, the barrier is unchanged from the Gō model. Intuitively, the addition of random non-native interactions tends to lower the free energy in general. For the free energy *barrier* to change, the relative stability of the unfolded and transition state must change. The only way this can happen upon increasing non-native interactions is if the densities of non-native polymer are different in the unfolded and transition states. Thus, effects on the barrier must come from the coupling of non-native density to the degree of nativeness.

We then showed this coupling is such that generically, the non-native density increases with increasing nativeness, so long as the protein is sufficiently large and flexible, and the non-native interactions are weak. Thus, the free-energy barrier lowers as non-native interaction strength is increased from zero.

The folding rate increases if the effect of ruggedness on the prefactor is weaker than the effect on the barrier. The prefactor is related to the reconfigurational diffusion time. Because a dynamic glass transition is expected in such systems, there will be a window for weak ruggedness within which the diffusion time, and thus the rate prefactor, is relatively unaffected as ruggedness is increased. The rate tends to initially increase before decreasing substantially, as shown in Figure 1 and expressed in Eq. (27). This is the main result of our work.

One might have some concern over the use of Q as a reaction coordinate in the analysis here. However, the regime that we considered, namely a Gō model perturbed by weak ruggedness, is precisely the regime where Q works best as a reaction coordinate. Indeed, simulations of some off-lattice model systems observe rate-enhancement.[7]

For wild-type proteins, which are very well-designed, the effect may be observable through calorimetric measurements on mutant sequences that still fold to the native structure. The width of the unfolded density of states, which is a function of b, may be inferred from calorimetry[13] and compared with the folding rate for that sequence.

This phenomenon provides a good example of how energy landscape theory can be applied to the physics of protein folding to reveal and explain a counterintuitive result.

ACKNOWLEDGMENTS

We thank Jose' Onuchic, Cecilia Clementi, and David Clayton for many enjoyable and insightful discussions, and Peter Wolynes for helpful discussions on the collapse transition. We are also grateful to Hugh Nymeyer for providing the off-lattice simulation data and constructive advice.

REFERENCES

1. Alm E, Baker D. Prediction of protein-folding mechanisms from free-energy landscapes derived from native structures. Proc Natl Acad Sci USA 1999;96:11305–11310.
2. Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG. Funnels, pathways and the energy landscape of protein folding. Proteins 1995;21:167–195.
3. Bryngelson JD, Wolynes PG. Intermediates and barrier crossing in a random energy model (with applications to protein folding). J Phys Chem 1989;93:6902–6915.
4. Bryngelson JD, Wolynes PG. A simple statistical field theory of heteropolymer collapse with applications to protein folding. Biopolymers 1990;30:177–188.
5. Clementi C, Nymeyer H, Onuchic JN. Topological and energetic factors: what determines the structural details of the transition state ensemble and en-route intermediates for protein folding? An investigation for small globular proteins. J Mol Biol 2000;298:937–953.
6. Clementi C, Jennings PA, Onuchic JN. How native state topology affects the folding of dihydrofolate reductase and interleukin-1beta. Proc Natl Acad Sci USA 2000;97:5871–5876.
7. Clementi C, Plotkin SS, Onuchic JN. The effect of non-native

contacts on protein folding rate and mechanism. 2001. In preparation.
8. Dill KA, Bromberg S, Yue K, Fiebig KM, Yee DP, Thomas PD, Chan HS. Principles of protein folding—a perspective from simple exact models. Protein Sci 1995;4:561–602.
9. Dobson CM, Sali A, Karplus M. Protein folding: a perspective from theory and experiment. Angew Chem Int Ed Engl 1998;37:868–893.
10. Doig AJ, Sternberg MJE. Side-chain conformational entropy in protein folding. Protein Sci 1995;4:2247–2251.
11. Finkelstein AV, Badretdinov AYa. Rate of protein folding near the point of thermodynamic equilibrium between the coil and the most stable chain fold. Folding Design 1997;2:115–121.
12. Galzitskaya V, Finkelstein AV. A theotetical search for folding/unfolding nuclei in three dimensional protein structures. Proc Natl Acad Sci USA 1999;96:11299–11304.
13. Freire E, Biltonen RL. Statistical mechanical deconvolution of thermal transitions in macromolecules. I. Theory and application to homogeneous systems. Biopolymers 1978;17:463–479.
14. Garel T, Orland H, Pitard E. Protein folding and heteropolymers. In: Young AP, editor. Spin glasses and random fields. River Edge, NJ: World Scientific; 1998.
15. Nobuhiro Gō. Theoretical studies of protein folding. Annu Rev Biophys Bioeng 1983;12:183–210.
16. Goldstein RA, Luthey-Schulten ZA, Wolynes PG. Optimal protein-folding codes from spin-glass theory. Proc Natl Acad Sci USA 1992;89:4918–4922.
17. Götze W. Aspects of structural glass transitions. In: Levesque D, Hanson JP, Zinn-Justin J, editors. Liquids, freezing and the glass transition. New York: Elsevier; 1991. p 287–504.
18. Guerois R, Serrano L. The SH3-fold family: experimental evidence and prediction of variations in the folding pathways. J Mol Biol 2000;304:967–982.
19. Honeycutt JD, Thirumalai D. Metastability of the folded states of globular proteins. Proc Natl Acad Sci USA 1990;87:3526–3529.
20. Kaya H, Chan HS. Polymer principles of protein calorimetric two-state cooperativity. Proteins 2000;40:637–661.
21. Kirkpatrick TR, Wolynes PG. Stable and metastable states in mean-field potts and structural glasses. Phys Rev B 1987;36:8552–8564.
22. Landau LD, Lifshitz EM. Statistical physics, 3rd ed. Oxford: Pergamon Press; 1980.
23. Li L, Mirny LA, Shakhnovich EI. Kinetics, thermodynamics and evolution of non-native interactions in a protein folding nucleus. Nat Struct Biol 2000;7:336–342.
24. Li MS, Cieplak M. Simple models of proteins with repulsive non-native contacts. Eur Phys J 2000;14:787–792.
25. Makhatadze GI, Privalov PL. On the entropy of protein folding. Protein Sci 1996;5:507–510.
26. Munoz V, Eaton WA. A simple model for calculating the kinetics of protein folding from three-dimensional structures. Proc Natl Acad Sci USA 1999;96:11311–11316.
27. Nemirovsky AM, Dudowicz J, Freed KF. Thermodynamics of a dense self-avoiding walk with contact interactions. J Stat Phys 1992;67:395.
28. Nymeyer H, Garcia AE, Onuchic JN. Folding funnels and frustration in off-lattice minimalist protein landscapes. Proc Natl Acad Sci USA 1998;95:5921–5928.
29. Onuchic JN, Luthey-Schulten Z, Wolynes PG. Theory of protein folding: the energy landscape perspective. Annu Rev Phys Chem 1997;48:545–600.
30. Onuchic JN, Wolynes PG, Luthey-Schulten Z, Socci ND. Toward an outline of the topography of a realistic protein-folding funnel. Proc Natl Acad Sci USA 1995;92:3626–3630.
31. Pande VS, Grosberg AYu, Tanaka T. Heteropolymer freezing and design: towards physical models of protein folding. Rev Mod Phys 2000;72:259–314.
32. Plaxco KW, Millett IS, Segel DJ, Doniach S, Baker D. Chain collapse can occur concomitantly with the rate-limiting step in protein folding. Nat Struct Biol 1999;6:554–556.
33. Plotkin SS, Wang J, Wolynes PG. Statistical mechanics of a correlated energy landscape model for protein folding funnels. J Chem Phys 1997;106:2932–2948.
34. Plotkin SS, Wolynes PG. Non-markovian configurational diffusion and reaction coordinates for protein folding. Phys Rev Lett 1998;80:5015–5018.
35. Plotkin SS, Onuchic JN. Investigation of routes and funnels in protein folding by free energy functional methods. Proc Natl Acad Sci USA 2000;97:6509–6514.
36. Plotkin SS, Onuchic JN. Understanding protein folding with energy landscape theory Part II: Quantitative aspects. Q Rev Biophys 2001. Forthcoming.
37. Prieto J, Wilmans M, Jiménez MA, Rico M, Serrano L. Non-native local interactions in protein folding and stability: introducing a helical temdency in the all β-sheet α-Spectrin SH3 domain. J Mol Biol 1997;268:760–778.
38. Riddle DS, Grantchasoua VP, Santiago JV, Alm E, Ruczinski I, Baker D, et al. Experiment and theory highlight role of native state topology in SH3 folding. Nat Struct Biol 1999;6:1016–1024.
39. Shakhnovich EI. Protein design: a perspective from simple tractable models. Folding Design 1998;3:R45–R58.
40. Shea JE, Nochomivitz YD, Guo Z, Brooks CL. Exploring the space of protein folding hamiltonians: the balance of forces in a minimalist β-barrel model. J Chem Phys 1998;109:2895–2903.
41. Shoemaker BA, Wang J, Wolynes PG. Structural correlations in protein folding funnels. Proc Natl Acad Sci USA 1997;94:777–782.
42. Shoemaker BA, Wang J, Wolynes PG. Exploring structures in protein folding funnels with free energy functinals: the transition state ensemble. J Mol Biol 1999;287:675–694.
43. Socci ND, Onuchic JN. Kinetic and thermodynamic analysis of proteinlike heteropolymers: Monte Carlo histogram technique. J Chem Phys 1995;103:4732–4744.
44. Socci ND, Onuchic JN, Wolynes PG. Diffusive dynamics of the reaction coordinate for protein folding funnels. J Chem Phys 1996;104:5860–5868.
45. Takada S, Portman JJ, Wolynes PG. An elementary mode coupling theory of random heteropolymer dynamics. Proc Natl Acad Sci USA 1997;94:23188–2321.
46. Ueda Y, Taketomi H, Nobuhiro Gō. Studies on protein folding, unfolding, and fluctuations by computer simulation. Int J Peptide Res 1975;7:445–459.
47. Veitshans T, Klimov D, Thirumalai D. Protein folding kinetics: timescales, pathways and energy landscapes in terms of sequence-dependent properties. Folding Design 1997;2:1–22.
48. Wang J, Plotkin SS, Wolynes PG. Configurational diffusion on a locally connected correlated energy landscape: application to finite, random heteropolymers. J Phys I France 1997;7:395–421.
49. Wolynes PG. Spin glass ideas and the protein folding problems. In: Stein DL, editor. Spin glasses and biology. Singapore: World Scientific; 1992. p 225–259.

APPENDIX A: CAVEATS DUE TO FINITE-SIZE, GENERIC ATTRACTION, AND STIFFNESS EFFECTS

The derivation leading to Eq. (11) assumed mean-field theory could be applied, that the protein could be treated as a bulk system, and that properties arising from chain connectivity would not alter the results arising from the energetics in the problem. In particular, we have assumed that the polymer persistence length or Kuhn length ℓ_K is much less than the length of a typical piece of disordered protein ℓ_0 near the barrier peak, so that a dangling piece of disordered polymer may interact with itself. If the protein under study is particularly stiff and/or small, the return length of polymer fragments may be comparable to the length of the disordered pieces, reducing the number of non-native interactions near the barrier peak relative to the number in the unfolded state. Then in Eq. (10) the density $\eta(Q^{\neq}) \lesssim \eta(0)$ and no reduction in barrier height with non-native interactions would be seen.

For the 27-mer lattice model $\ell_K \approx 3$–4; these models are relatively stiff compared to their total length. For typical off-lattice models on the other hand, $\ell_K \approx 3$, but they are considerably longer, for example, for SH3, $N = 57$. If all the non-native polymer is in one strand, $\ell_0 \approx N/2$ at the barrier peak. Then $\ell_K/\ell_0 \approx 0.26$ for the lattice model, and $\ell_K/\ell_0 \approx 0.11$ for the SH3 off-lattice model. If the non-native

polymer is distributed among a number of disordered strands that can dress the native core, roughly $(1/6) \times N^{2/3}$,[11] then $\ell_K/\ell_0 \approx 0.40$ for the lattice model, and $\ell_K/\ell_0 \approx 0.26$ for the off-lattice model.

Neither of these numbers is very small, indicating that the collapse transitions are quite rounded, and the effect on folding rate will be mild if it exists. In fact, the discrepancy of ℓ_K/ℓ_0 between the off-lattice and on-lattice models, although fairly small here, leads to different behavior of the non-native density $\eta(Q)$, as shown in Figure 4. In the lattice system $\eta(Q^{\neq}) \lesssim \eta(Q_U)$, but in the off-lattice system $\eta(Q^{\neq}) \gtrsim \eta(Q_U)$. Hence, we anticipate the rate-enhancement effects will be seen in off-lattice models but probably not in at least the shorter on-lattice models.[7]

Real proteins may tend to have some net homopolymer attraction inducing generic collapse. This decreases the change in density upon folding and would further attenuate any rate enhancement effect present. However, at least some proteins are sufficiently stable that collapse and folding are concomitant.[32] Moreover, Gō models, for which collapse and folding are concomitant by construction, capture at least some of the essential aspects of folding mechanism.[1,5,6,12,26,41,42]

Collapse accompanies folding when the folding transition temperature T_F given by Eq. (8) is comparable to the collapse temperature T_θ in Eq. (15). For weak ruggedness,

$$\frac{T_F}{T_\theta} \approx \frac{\epsilon}{as_o}\left(1 - \frac{s_o}{2z}\frac{b^2}{\epsilon^2} - \frac{1}{2z}\frac{b^2}{a^2}\right). \tag{30}$$

So the effects of generic collapse are not important in the problem as long as $\epsilon \gtrsim as_o$, to the first approximation.

Several additional features may affect the folding rate. In finite-sized systems, the unfolded state tends to have partial order. Moreover, its position may drift, along with that of the transition state, as non-native variance is increased. This modifies the barrier height. The density $\eta(Q)$ in Figure 4 is taken from Gō models and so is exact in the limit $b \to 0$. However, for nonzero b, the density $\eta(Q)$ may begin to alter in structure, as non-native interactions induce collapse and transient traps. Accounting for these effects will modify the folding rate, but should not alter the general trend of Figure 1. A nonzero mean repulsive or attractive non-native interaction strength, investigated in lattice models (e.g., Refs. 8, 24, and 43), may be straightforwardly incorporated into the theory developed here. Such a parameter would couple stability gap to mean homopolymer attraction in the theory.