Determination of Barrier Heights and Prefactors from Protein Folding Rate Data

S. S. Plotkin

Department of Physics and Astronomy, University of British Columbia, Vancouver, British Columbia, Canada

ABSTRACT   We determine both barrier heights and prefactors for protein folding by applying constraints determined from experimental rate measurements to a Kramers theory for folding rate. The theoretical values are required to match the experimental values at two conditions of temperature and denaturant that induce the same stability. Several expressions for the prefactor in the Kramers rate equation are examined: a random energy approximation, a correlated energy approximation, and an approximation using a single Arrhenius activation energy. Barriers and prefactors are generally found to be large as a result of implementing this recipe, i.e., the folding landscape is cooperative and smooth. Interestingly, a prefactor with a single Arrhenius activation energy admits no formal solution.

INTRODUCTION

In contrast to many exothermic reactions in organic chemistry, the log protein folding rate displays a significant linear trend with the relative stability of the product and reactant (folded and unfolded states; Fersht, 1999). This indicates a late transition state in the language of Hammond's postulate, and the slope of the log rate versus stability line quantifies the degree of native structural information in the transition state.

Native stability may be modified by adjusting temperature T or denaturant concentration c. Many proteins show linearity over the majority of the branches of their Chevron plot, implying a linear dependence of folding and unfolding barriers on denaturant concentration c (Jackson and Fersht, 1991),

$$\Delta G_{F\ddagger}(T, c) \equiv G_{\ddagger}(T, c) - G_F(T, c) = \Delta G_{F\ddagger}(T, 0) - m_{F\ddagger}c, \quad (1a)$$

$$\Delta G_{U\ddagger}(T, c) \equiv G_{\ddagger}(T, c) - G_U(T, c) = \Delta G_{U\ddagger}(T, 0) + m_{U\ddagger}c, \quad (1b)$$

with $m_{F\ddagger} > 0$ and $m_{U\ddagger} > 0$.

Subtracting Eq. 1b from Eq. 1a, and defining $\Delta G \equiv \Delta G_{FU} = G_U - G_F$ and $m = m_{F\ddagger} + m_{U\ddagger}$, we have that

$$\Delta G(T, c) = \Delta G(T, 0) - mc. \quad (2)$$

For two-state folders, the kinetically determined m above equals, to good approximation, the thermodynamically determined m-value from relative stabilities.

Applying Kramers rate theory, the log forward folding rate is given by

$$\ln k_F(T, c) = \ln k_o(T, c) - \Delta G_{U\ddagger}(T, c)/T = \ln k_o(T, c) \\ - (\Delta G_{U\ddagger}(T, 0) + m_{U\ddagger}c)/T. \quad (3)$$

Eliminating c from Eqs. 2 and 3 gives

$$\ln k_F(T, c) - \frac{m_{U\ddagger}}{m} \frac{\Delta G(T, c)}{T} = \ln k_o(T, c) \\ - \frac{1}{T}\Big(\Delta G_{U\ddagger}(T, 0) + \frac{m_{U\ddagger}}{m}\Delta G(T, 0)\Big), \quad (4)$$

where the left-hand side of Eq. 4 depends on both (T, c), but the function on the right-hand side depends on c only through the prefactor. Empirically it was observed by Scalley and Baker (1997) that for the proteins CspB and protein L, the data for various c collapse onto a single curve when the left-hand side is plotted versus $1/T$. This indicates that the right-hand side is a function of temperature alone and so $\ln k_o(T, c) \approx \ln k_o(T)$. Denaturant concentration does not have a significant effect on the rate at which the system escapes from local traps (at least for those proteins studied). We make this assumption here as well.

Because the prefactor is independent of c, the change in log folding rate with denaturant is directly proportional to the change in barrier with denaturant,

$$\delta \ln k_F \equiv \ln k_F(T, c) - \ln k_F(T, 0) \\ = -(\Delta G_{U\ddagger}(T, c) - \Delta G_{U\ddagger}(T, 0))/T = -\delta\Delta G_{U\ddagger}/T, (5)$$

which, together with Eq. 4, gives

$$\delta\Delta G_{U\ddagger} = -(m_{U\ddagger}/m)\delta\Delta G, \quad (6)$$

$$\delta \ln k_F = (m_{\ddagger}/m)(\delta\Delta G/T). \quad (7)$$

This quantifies the assertion above that log folding rates depend linearly on the relative stability of the products. If we let $m_{U\ddagger}/m \equiv Q^{\neq}$, Eq. 6 can be rewritten as

$$\delta G_{\ddagger} = Q^{\neq}\delta G_F + (1 - Q^{\neq})\delta G_U, \quad (8)$$

which is the commonly used linear free energy relation (Bryngelson et al., 1995).

Inspection of rate-stability isotherms for several different proteins—cytochrome C (cyt-C; Mines et al., 1996), protein

L (Scalley and Baker, 1997), cspB (Schindler and Schmid, 1996), N-terminal protein L9 (Kuhlman et al., 1998), and S6 (Otzen and Oliveberg, 2004)—shows linearity over ranges up to ≈ 25 kJ $\times$ mol$^{-1} \approx 10\,k_BT$, indicating large and robust folding barriers, which are substantially larger than the folding barriers seen in many simulations (for example, see Fig. 1).

At a higher temperature, the log rate versus stability curve is still linear, with approximately the same slope, indicating the nativeness of the transition state, in terms of solvent exposure, is not significantly changed (Fig. 1). However, the rates are higher, presumably due to two effects:

1. The prefactor increases at higher temperature (since activated escape from traps is further facilitated, and solvent viscosity is reduced).
2. The thermodynamic weight of the entropic component to the barrier (which includes contributions from the solvent) increases as well, which may decrease the barrier height.

METHODOLOGY

In what follows, we apply Kramers rate theory together with energy landscape ideas to extract barrier heights and prefactors from experimental rate data.

The temperature-dependence of the stability is given by the Gibbs-Helmholtz expression (Fersht, 1999; Jackson and Fersht, 1991) as

$$\Delta G(T, c) = \Delta H - T\Delta S + \Delta C_P(T - T_o - T\ln(T/T_o)). \quad (9)$$

Then at equal stabilities $\Delta G(T_o, c_o) = \Delta G(T, c)$,

$$\Delta C_P(T - T_o - T\ln(T/T_o)) = (T - T_o)\Delta S + m(c - c_o). \quad (10)$$

For two-state folders, the heat-capacity ratio $\Delta C_{PU\ddagger}/\Delta C_P$ is approximately equal to m-value ratio $m_{U\ddagger}/m$, giving the fractional solvent accessibility of the transition state. We assume this equality here as well, which gives for Eq. 10,
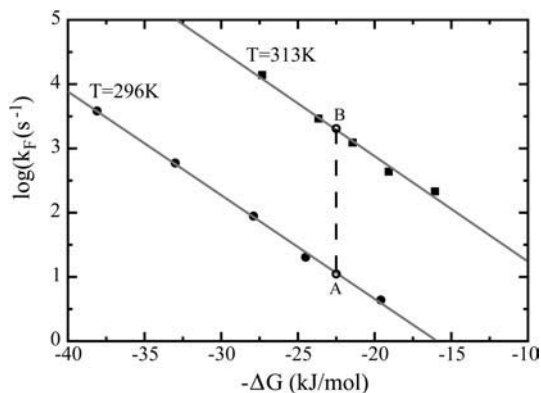

FIGURE 1 Logarithm of the rate versus (minus) native stability for horse Cytochrome C, at two temperatures. The plots are well fit by straight line functions that are used in the analysis of the text. Adapted from Mines et al. (1996).

$$\Delta C_{PU\ddagger}(T - T_o - T\ln(T/T_o)) = -\frac{m_{U\ddagger}}{m}\Delta S(T - T_o)$$
$$- m_{U\ddagger}(c - c_o). \quad (11)$$

Inserting Eq. 11 into Gibbs-Helmholtz expressions for the barrier heights $\Delta G_{U\ddagger}$ at (T, c) and (T_o, c_o) gives the change in barrier height at fixed stability,

$$[\Delta G_{U\ddagger}(T, c) - \Delta G_{U\ddagger}(T_o, c_o)]_{\Delta G(T,c)=\Delta G(T_o,c_o)} \equiv \delta'\Delta G_{U\ddagger}$$
$$= -(T - T_o)\left(\Delta S_{U\ddagger} + \frac{\Delta m_{U\ddagger}}{m}\Delta S\right), \quad (12)$$

which is independent of c and depends only on the temperature difference between the two fixed-stability states (and thermodynamic parameters). This equation applies to points A and B in Fig. 1, for example.

For changes in temperature of a few degrees, the change in barrier height $\delta'\Delta G_{U\ddagger}$ is only a few percent of the total barrier height, when the rates versus temperature and denaturant are fit to a model to extract thermodynamic parameters, as in Kuhlman et al. (1998), Otzen and Oliveberg (2004), Scalley and Baker (1997), and Schindler and Schmid (1996). We used Eq. 12 for the change in barrier height when thermodynamic data were available. For the case of cyt-C we set $\delta'\Delta G = 0$.

The rates for pairs of states at the same stability ΔG are given from Eq. 3 as

$$\ln k_F(T_o, c_o) = \ln k_o(T_o) - \Delta G_{U\ddagger}(T_o, c_o)/T_o, \quad (13a)$$

$$\ln k_F(T, c) = \ln k_o(T) - \Delta G_{U\ddagger}(T_o, c_o)/T - \delta'\Delta G_{U\ddagger}/T. \quad (13b)$$

Random energy model for the temperature-dependent prefactor

At the mean field level for a landscape of uncorrelated states (random energy model or REM), the temperature-dependence of the prefactor in Eq. 3 is super-Arrhenius (Bryngelson and Wolynes, 1989; Onuchic et al., 1997). Moreover, the prefactor goes as the reciprocal of the viscous friction coefficient (Hanggi et al., 1990; Klimov and Thirumalai, 1997; Socci et al., 1996), so the log prefactors at (T_o, c_o) and (T, c) may be written as

$$\ln k_o(T_o) = \ln k_{oo} - \Delta^2/2T_o^2, \quad (14a)$$

$$\ln k_o(T) = \ln k_{oo} - \Delta^2/2T^2 + \ln(\eta(T_o)/\eta(T)). \quad (14b)$$

To compare rate theories with experimental data we must introduce a fundamental timescale or rate constant k_{oo}, which is then modified by barriers representing the ruggedness of the energy landscape. Rates for short loop closure are $\sim 2 \times 10^7$ s^{-1} (Lapidus et al., 2000), comparable to helix formation rates of $\sim 10^7$ s^{-1}, and somewhat faster than rates of hairpin formation $\sim 10^6$ s^{-1} (Eaton et al., 2000). Prefactors obtained from plots of experimental rate versus powers of chain length are of order μs (Li et al., 2004); however, these implicitly include any effects due to ruggedness. We take 10^7 s^{-1} as an estimate of the fastest local rate. Since $\sim$10–100 loops and/or secondary structural elements exist in a protein, we then take $k_{oo} = 10^9$ s^{-1}. This estimate for k_{oo} may appear somewhat large; we will see later that smaller estimates for k_{oo} give smaller estimates for inferred folding barriers. We use the known temperature dependence of the viscosity in water (CRC, 2003). The quantity Δ^2 measures the ruggedness of the energy landscape. It may be eliminated from Eqs. 14a and 14b to give an equation relating the prefactors as

$$\ln k_o(T) = \left(1 - \frac{T_o^2}{T^2}\right)\ln k_{oo} + \frac{T_o^2}{T^2}\ln k_o(T_o) + \ln\left(\frac{\eta(T_o)}{\eta(T)}\right). \quad (15)$$

Equations 13a, 13b, and 15 constitute a system of three linear equations for three unknowns: $\Delta G_{U\ddagger}(T_o, c_o)$, $\ln k_o(T_o)$, and $\ln k_o(T)$, which can be solved analytically at any given stability, from linear fits to the log rate-stability data.

RESULTS

The results of applying the method are shown in Fig. 2, for the data in Fig. 1, ranging from the stability of wild-type at 296 K (−74 kJ/mol) to zero stability at the transition midpoint. Barrier heights are plotted in units of kJ/mol; rates in prefactors are in units of s^{-1}.

We can see several things from this plot. The barrier heights at the transition midpoint are large, compared to values obtained from simulation models as well as theories with pair interaction potentials. If the linear relation in Eq. 6 held until the transition midpoint, the barrier would be ∼30 kJ/mol plus whatever the barrier was at conditions of zero denaturant.

The slope $\delta\Delta G_{U\ddagger}/\delta\Delta G \approx 0.8$ is also larger than its empirical value of $m_{U\ddagger}/m \approx 0.4$ (Mines et al., 1996), thus the barriers vanish at weaker stabilities than the wild-type protein. This indicates a breakdown in the validity of the theory at higher stabilities (larger ΔG).

There are two parameters in the theory for which we have put in approximate values: the value of the attempt frequency $k_{oo} = 10^9$ s^{-1}, and the value of $\delta'\Delta G_{U\ddagger}$, which we have set to zero for cytochrome C in the absence of an empirically determined value. Increasing k_{oo} or decreasing $\delta'\Delta G_{U\ddagger}$ raises barriers, but does not change the slope $\delta\Delta G_{U\ddagger}/\delta\Delta G$. The value of $-\Delta G$ where the barrier vanishes linearly decreases as $\delta'\Delta G_{U\ddagger}$ is decreased below zero, with the barrier vanishing at the stability of the wild-type when
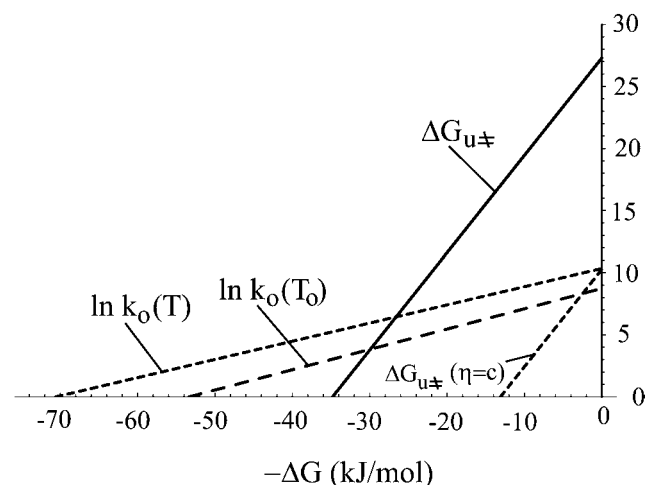

FIGURE 2 Barrier height $\Delta G_{U\ddagger}$ and prefactors k_o at two temperatures, as obtained from the REM approximation (see text), are plotted as a function of minus stability for cyt-C. The wild-type protein has a stability of $\Delta G \approx 74$ kJ/mol. Numerical values are given in Table 1. Prefactor attempt rates are in s^{-1}, and barrier heights are in kJ/mol. The short dashed line gives the barrier for a temperature-independent solvent viscosity. All logarithms are natural (base e).

$\delta'\Delta G_{U\ddagger}$ is ∼−1.6 kJ/mol. This is not an unreasonable number compared to experimental numbers for other proteins (see below); however, it is somewhat disconcerting that barrier heights are such a strong function of the barrier change $\delta'\Delta G_{U\ddagger}$. We will see later that this sensitivity is not present when a correlated landscape model is used for the prefactor.

Fig. 2 also shows that at least for the REM approximation it is important to account for changes in the viscosity of the solution with temperature, as the barrier substantially decreases when the viscosity is held constant versus temperature.

Equations 14a or 14b may now be solved for Δ, giving a number ≈15 kJ/mol, that only weakly depends on stability ΔG or barrier change $\delta'\Delta G_{U\ddagger}$. Estimating the chain conformational entropy as ∼100 k_B (D'Aquino et al., 1996; Leach et al., 1966), we can give an estimate for the glass temperature T_G for this system,

$$T_G = \Delta/(2S_o/k_B)^{1/2}, \tag{16}$$

which is also a fairly robust number as a function of stability or barrier change, as shown in Fig. 3. At the stability of wild-type cyt-C, $T_G \approx 150$ K, giving $T/T_G \approx 2.0$ at 296 K.

Correlated landscape model for the temperature-dependent prefactor

Many of the problems of the REM approximation are resolved by accounting for pair correlations between states in the expression for the prefactor. Below a critical temperature T_A on a correlated landscape, dynamics are activated, and the rate prefactor increases as temperature is raised (Plotkin and Onuchic, 2002a,b; Wang et al., 1997). The expressions for the rate prefactors at T_o and T become

$$\ln k_o(T_o) = \ln k_{oo} - (S^{\neq}/2)\left(\alpha - \beta(1 - T_G/T_o)^2\right), \tag{17a}$$

$$\ln k_o(T) = \ln k_{oo} - (S^{\neq}/2)\left(\alpha - \beta(1 - T_G/T)^2\right) + \ln(\eta(T_o)/\eta(T)). \tag{17b}$$

Here $S^{\neq}$ is the chain entropy at the transition state, and α and β are parameters measuring the mismatch between entropy and energy giving the typical free energy barrier governing trap escape. The values for a bulk polymer $\alpha \approx 0.5$, $\beta \approx 1.8$ are used below (Plotkin and Onuchic, 2002a,b; Wang et al., 1997). The temperature T_G was adjusted to the value that reproduced the experimentally determined slope of barriers versus stability, $m_{U\ddagger}/m$. In Table 1 this number is compared to the value of T_G that emerges from the REM analysis. A mismatch of these two values may indicate a breakdown of the REM approximation for states in determining prefactors, i.e., a breakdown in the validity of Eqs. 14a and 14b. For cyt-C the value of T_G giving the correct slope is ∼1.2 kJ/mol, versus 1.0 kJ/mol from the REM analysis.

The entropy $S^{\neq}$ may be eliminated from Eqs. 17a and 17b, giving an equation that relates the prefactors, which replaces Eq. 15,
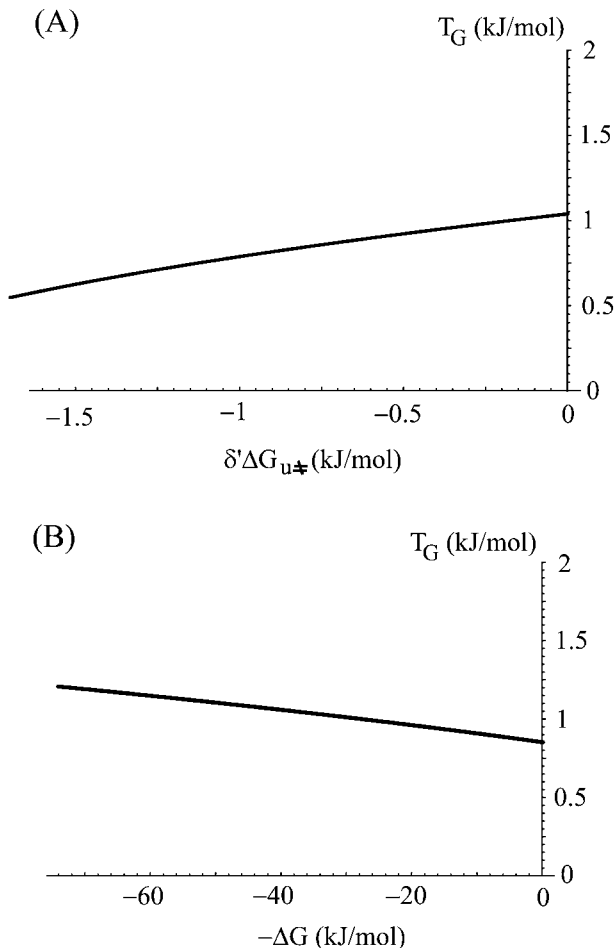
FIGURE 3 (A) The temperature T_G that emerges from the REM analysis for cyt-C (see text and Eq. 16) varies only moderately with barrier height change at constant stability, $\delta'\Delta G_{U\ddagger}$ (the value of which is not known for this protein). For this plot the stability is set to midway between zero and the stability of the wild-type (37 kJ/mol). (B) T_G also changes little as native stability ΔG is varied (for this plot $\delta'\Delta G_{U\ddagger} = 0$).

$$[\ln k_{oo} - \ln k_o(T) + \ln(\eta(T_o)/\eta(T))][\alpha - \beta(1 - T_G/T_o)^2]$$
$$= [\ln k_{oo} - \ln k_o(T_o)][\alpha - \beta(1 - T_G/T)^2]. \qquad (18)$$

Equations 13a, 13b, and 18 again define a system of three linear equations for three unknowns: $\Delta G_{U\ddagger}(T_o, c_o)$, $\ln k_o(T_o)$, and $\ln k_o(T)$, which may be solved analytically. The results are shown in Fig. 4.

We see that both barriers and prefactors are larger than the corresponding REM values for cyt-C and the analysis for other proteins yields quite large numbers in general (compare to Table 1 for numbers). The barriers at the transition midpoint are ~22 $k_B T_{300K}$, and prefactors are almost unactivated.

The REM value of T_G resulted from approximating a value of 100 k_B for the chain entropy S_o, so it is feasible that this estimate for the REM T_G could differ from the T_G that gives the correct $m_{U\ddagger}/m$. The parameters α and β could, in prin-

ciple, have been adjusted to best match the experimental slope; however, it can be shown that this results in the same solution of Eqs. 13a, 13b, and 18 as that determined by varying T_G.

In contrast to the REM approximation, the effects of the temperature dependence of viscosity were not significant here (Fig. 4). Nor were there any significant effects due to barrier height difference—as $\delta'\Delta G_{U\ddagger}$ changed from -2 kJ/mol to 0 kJ/mol, the barrier changed by $<2\%$. The effects due to T_G are modest as well: over the range of T_G values in Fig. 3 B, the barrier height changed by $<15\%$. Lastly, the prefactors of the correlated landscape model are nearly constant over the range of experimental stabilities (Fig. 4), consistent with empirical observations (see the comments below Eq. 4).

Equation 17a or 17b may now be solved for $S^{\neq}$ as a check, giving $S^{\neq} \approx 40$ k_B, or $\sim40\%$ of the unfolded chain entropy assumed in finding the REM T_G. Alternatively, we can estimate the unfolded entropy S_o from the value of $S^{\neq}$ as $S^{\neq} \approx (1 - m_{U\ddagger}/m)S_o$, then Eq. 16 gives $\Delta \approx 14$ kJ/mol. Since the variances of individual residues add to give Δ^2, $\Delta^2 \approx N(1 - m_{U\ddagger}/m)b^2$, where b is a non-native energy scale per residue, here as ≈0.7 $k_B T_{300}$.

Fig. 5 shows that the inferred barriers and prefactors increase as the value of the bare reconfiguration rate k_{oo} increases. The prefactor $\ln k_o(T_o)$ closely follows the bare reconfiguration rate $\ln k_{oo}$; i.e., they are approximately equal. The barriers at the transition midpoint $\Delta G_{U\ddagger}^o$ and at the stability of the wild-type protein $\Delta G_{U\ddagger}^{(wt)}$ increase linearly, as $\sim2k_B T_o \ln k_{oo}$.

In the REM analysis there is an intermediate regime where the prefactor has a more complex temperature dependence than Eq. 14a. We do not describe this regime in detail since it is obtained from Eqs. 17a and 17b in the limit that $\alpha \to 1$, $\beta \to 2$, $S^{\neq} \to S_o$. Values obtained tended to be bracketed by the REM and correlated models.

For NTL9, the solution of the REM gave a T_G that monotonically decreased from a value of 0.4 at the stability of the wild-type protein, to zero at a stability of ~11 kJ/mol. Similarly, the prefactor monotonically increases from 10^8 s^{-1} at the stability of the wild-type to 10^{10} s^{-1} at zero stability. We note that these problems are not present if the stability difference $\delta'\Delta G_{U\ddagger}$ is set to zero, if the prefactor is two or more orders-of-magnitude slower, or if the temperature dependence of the viscosity is neglected. We take this sensitivity as a shortcoming of the procedure of rigorously demanding that the landscape theory fit to a limited subset of the experimental data. In this sense, a best (but not exact) fit to experimental rate surfaces as a function of both T and c as in Kuhlman et al. (1998), Otzen and Oliveberg (2004), Scalley and Baker (1997), and Schindler and Schmid (1996), is likely to give more accurate numbers. Likewise in the correlated model for NTL9, the prefactor increased from $\sim10^8$ s^{-1} at the stability of the wild-type to unphysical values at zero stability. A similar situation exists in the REM

TABLE 1 Thermodynamic and kinetic parameters for proteins studied

Proteins*	T_o	T	$\Delta G^{\mathrm{wt}\dagger}$	$\delta' \Delta G_{\mathrm{U}\ddagger}$	Correlated model				Random energy model			
					$\Delta G_{\mathrm{U}\ddagger}^{0\ \dagger\dagger}$	$\Delta G_{\mathrm{U}\ddagger}^{\mathrm{wt}\ \S}$	$k_0(T_0)^{\dagger\dagger}$	$T_{\mathrm{G}}^{(\mathrm{m})\P}$	$\Delta G_{\mathrm{U}\ddagger}^{0\ \dagger\dagger}$	$\Delta G_{\mathrm{U}\ddagger}^{\mathrm{wt}\ \S}$	$k_0(T_0)^{\dagger\dagger}$	$T_{\mathrm{G}}^{\mathrm{REM}\|}$
cyt-C	2.46	2.60	−74	0	56	27	5×10^8	1.2	27	0	6×10^3	1.0
NTL9	2.48	2.59	−19	−1.0	47	35	(6×10^9)	1.7	50	33	(10^{10})	0 (0.4)**
S6	2.48	2.56	−31	−1.4	61	39	9×10^8	1.2	78	21	(10^{12})	0 (0.7)**
PTL	2.34	2.43	−22	−0.8	58	45	1×10^9	1.1	56	27	3×10^8	0.5
CspB	2.38	2.44	−9	−0.8	7	0	10^2	1.9**	24	20	2×10^5	0.7

*Sources for experimental data: cyt-C (Mines et al., 1996), NTL9 (Kuhlman et al., 1998), S6 (Otzen and Oliveberg, 2004), PTL (Scalley and Baker, 1997), and cspB (Schindler and Schmid, 1996). All temperatures and energies are in kJ/mol. All rates are in s^{-1}.
$\dagger$Stability of the wild-type protein.
$\dagger\dagger$At the transition midpoint where $\Delta G = 0$.
$\S$At the stability of the wild-type protein, where $c = 0$. If the barrier vanished at stabilities below the wild-type, the barrier value was simply taken as zero.
$\P$Value of T_{G} that gives a slope of barrier height versus stability equivalent to the experimental value of $m_{\mathrm{U}\ddagger}/m$.
$\|$Value of T_{G} using the REM approximation for rates, taken at a stability of approximately one-half of the wild-type protein.
**See text for explanation and comments.

recipe for protein S6; however, it is resolved in the correlated landscape model for that protein.

CspB showed some difficulties that arose from its unusually late transition state ($m_{\mathrm{U}\ddagger}/m \approx 0.9$) (Perl et al., 2002). The parameter T_{G} in the correlated model could not be adjusted to reproduce the high slope of barrier versus stability, without giving negative barriers. Again this may be an artifact of the exact fitting method mentioned above, i.e., more experimental data may also be needed to obtain more accurate numbers, or it may indicate that a simple mean field prefactor does not fully adequately describe the folding dynamics of this protein. In this case we took the temperature $T_{\mathrm{G}} = 1.81$ kJ/mol that induced the barrier to vanish at the stability of the wild-type protein (while acknowledging that other fits give large barriers (Perl et al., 2002)). This has

a steep barrier-stability curve, with slope $m_{\mathrm{U}\ddagger}/m = 0.8$ (as opposed to 0.9 observed empirically), very small barrier (7 kJ/mol at zero stability), and rugged landscape with very slow prefactor ($\sim 10^2$ s^{-1}). Such small barriers are consistent with estimates taken from simulations using C_α-models (Shea and Brooks, 2001).

The Arrhenius model generally admits no solution

A model often proposed for the prefactor assumes an Arrhenius temperature-dependence with single activation energy E_{A}, so that Eqs. 14a and 14b are replaced by

$$\ln k_{\mathrm{o}}(T_{\mathrm{o}}) = \ln k_{\mathrm{oo}} - E_{\mathrm{A}}/T_{\mathrm{o}} \tag{19a}$$

$$\ln k_{\mathrm{o}}(T) = \ln k_{\mathrm{oo}} - E_{\mathrm{A}}/T + \ln(\eta(T_{\mathrm{o}})/\eta(T)), \tag{19b}$$

from which E_{A} may be eliminated, yielding

$$\ln k_{\mathrm{o}}(T) = (1 - T_{\mathrm{o}}/T)\ln k_{\mathrm{oo}} + (T_{\mathrm{o}}/T)\ln k_{\mathrm{o}}(T_{\mathrm{o}}) \\ + \ln(\eta(T_{\mathrm{o}})/\eta(T)). \tag{20}$$
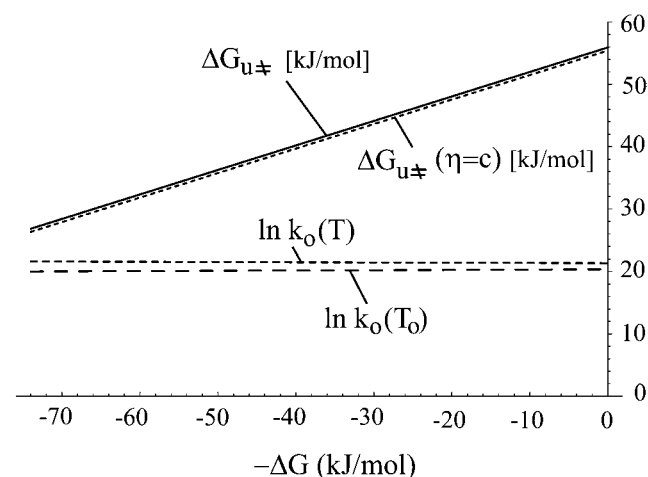
FIGURE 4 Barrier heights and prefactors as obtained from the correlated landscape analysis (see text), plotted as a function of minus native stability for h-cyt-C. Numerical values are given in Table 1. Prefactor attempt rates are in s^{-1}, and barrier heights are in kJ/mol. The dotted line gives the barrier for a temperature-independent solvent viscosity. Note prefactors are approximately constant (as is physically reasonable) and solvent viscosity plays a minor role.
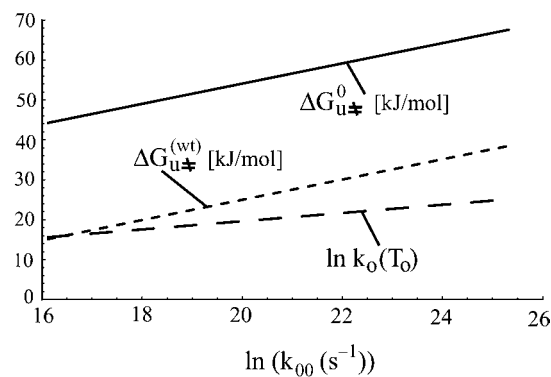
FIGURE 5 Barrier heights and prefactors extracted from the recipe for the correlated energy landscape (see text) increase as the bare reconfiguration rate (appearing in Eqs. 17a and 17b) increases. The increase is linear. $\Delta G_{\mathrm{U}\ddagger}^{\mathrm{o}}$ is the barrier at the transition midpoint, $\Delta G_{\mathrm{U}\ddagger}^{(\mathrm{wt})}$ is the barrier at the stability of the wild-type protein, and $k_{\mathrm{o}}(T_{\mathrm{o}})$ is the prefactor at temperature T_{o} in s^{-1}.

This equation, relating the prefactors together with Eqs. 13a and 13b, constitutes the new system of equations to be solved.

Eliminating $\Delta G_{U\ddagger}$ from Eqs. 13a and 13b gives another equation relating the prefactors:

$$\ln k_o(T) = \ln k_F(T, c) - (T/T_o)\ln k_F(T_o, c_o) + \delta'\Delta G_{U\ddagger}/T$$
$$+ (T_o/T)\ln k_o(T_o). \tag{21}$$

Equations 21 and 20 both have $\ln k_o(T)$ on the left-hand side and $(T_o/T)\ln k_o(T_o)$ on the right. Subtracting them gives an equation that is independent of any variable to be solved for

$$\ln k_F(T, c) - (T/T_o)\ln k_F(T_o, c_o) + \delta'\Delta G_{U\ddagger}/T$$
$$= (1 - T_o/T)\ln K_{\infty} + \ln(\eta(T_o)/\eta(T)), \tag{22}$$

which cannot be true in general, in particular because the left-hand side depends on c and the right-hand side does not.

A geometric analog may be helpful in understanding the situation. The solution to three equations in three variables is equivalent to finding the point where three planes intersect. Letting

$$x_1 = \ln k_o(T_o)$$
$$x_2 = \ln k_o(T)$$
$$x_3 = \Delta G_{U\ddagger},$$

Eqs. 13a, 13b, and 20 may be recast as

$$x_2 - (T_o/T)x_1 = A \tag{23a}$$
$$x_2 - (T_o/T)x_1 = B \tag{23b}$$
$$x_1 - (1/T_o)x_3 = C, \tag{23c}$$

where

$$A = (1 - T_o/T)\ln k_{\infty} + \ln(\eta(T_o)/\eta(T)),$$
$$B = \ln k_F(T_o, c_o) + (T_o/T)\ln k_F(T, c) + \delta'\Delta G_{U\ddagger}/T,$$
$$C = \ln k_F(T_o, c_o).$$

Since $A \neq B$ in general, Eqs. 23a and 23b depict two parallel planes. Thus there is no point of intersection and the system of equations is ill-posed. For the special case of $A = B$ there is a whole family of solutions consistent with the rate equations, but as mentioned above this scenario can only hold under very special circumstances.

CONCLUSIONS AND DISCUSSION

We have proposed here a method of testing energy landscape theory by mapping Kramers rate theory, with prefactors given from the statistics of energies of states, to experimental data on protein folding rates. We considered three models for the prefactor here: one where ruggedness is treated with a random energy approximation; one where correlations are taken into account; and an Arrhenius model with a single barrier governing reconfiguration times.

The numerical values of the barriers obtained from the above recipes should be taken with a grain of salt; however, it consistently emerged that folding barriers were large (except for CspB): the average barrier at the transition midpoint for the REM analysis is $\sim 19\ k_BT$, and the corresponding barrier in the correlated model is $\sim 18\ k_BT$. If CspB is omitted, the barriers are $21\ k_BT$ and $22\ k_BT$, respectively. Wild-type S6, a protein known to fold very cooperatively (Lindberg et al., 2002), had the highest barriers.

With the exception of CspB, the prefactors in the correlated model tended to be quite high—approximately the bare reconfiguration rate for the whole protein ($10^9\ s^{-1}$). The folding barrier obtained from the recipe decreases as estimates for the bare reconfiguration rate decrease (Fig. 5). The prefactors from the REM recipe varied considerably.

All of the proteins analyzed here are considered two-state folders, so we would expect a Kramers theory to describe them. In lower temperature regimes the distribution of first passage times may be more relevant to study (Plotkin and Onuchic, 2002b; Zhou et al., 2003).

We found that in practice it was quite important to have accurate fits for the empirical rate-stability curves. For example, as temperature increased, the slope of the log rate versus stability curve had to remain approximately constant or tend to increase, to obtain reasonable solutions of the rate equations. Otherwise we found an unphysical situation where barriers did not increase as stability decreased. This sensitivity to the experimental data may favor a less stringent fit to the experimental constraints.

In fact, reflection on the procedure raises a general issue on the rigorous application of experimental constraints to energy landscape theory. For example, if we were to add data at a third temperature T_1, two new equations would be introduced according to the recipe—one Kramers rate equation and one landscape equation for the prefactor, but only one new variable is introduced—the prefactor $\ln k_o(T_1)$. The system becomes overdetermined. Demanding equality rather than a best fit at several temperatures becomes too stringent a constraint on the theory, as long as the parameters in the theory (e.g., Δ^2 or E_A) are fixed. The more temperatures used, the more variables must be introduced into the theory, or the parameters must themselves become temperature-dependent. Nevertheless, the fact that the Arrhenius activation model fails in general to provide a solution for even two temperatures (two data points) should probably be seen as evidence against its strict applicability.

A perhaps more viable method would be to fit several temperatures with functional forms such as Eqs. 14a, 17a, or 19a to extract parameters such as Δ^2 and E_A. The difficulty in previous fits to data has been in the separation of E_A and the activation enthalpy $\Delta H_{U\ddagger}$ (Scalley and Baker, 1997). One can ask which temperature dependence (E_A/T or Δ^2/T^2) gives the best fit to the data, but there is not yet enough accurate data to distinguish between the two scenarios (Kuhlman et al., 1998; Scalley and Baker, 1997) by this method. How-

ever, the Arrhenius model becomes severely restricted by applying experimental constraints rigorously at two temperatures and denaturant concentrations, at the same stability. Because the activation energy in the prefactor can be absorbed into the enthalpic part of the barrier, and only the entropic part of the barrier is relevant in determining rate differences at fixed stability (by Eq. 12), the activation energy becomes irrelevant, and the difference in rates must then be due to quantities independent of denaturant concentration (entropic part of the barrier, temperature-dependent viscosity, etc.). All rate-stability curves for a given protein must be exactly parallel in the Arrhenius model—a situation not observed empirically.

Topological features of the native structure have been neglected in the rate theory. Including polymer physics into the theoretical model (Plotkin and Onuchic, 2000; Portman et al., 2001; Shoemaker et al., 1999) may also eliminate some of the sensitivity of the theoretically derived values in Table 1 on the experimental data.

Other methods have been used to estimate barrier heights. Adding a three-body contribution to a pairwise-interacting energy function to give best agreement with experimental ϕ-values, a barrier height for protein L of $\sim$16 kJ/mol was obtained (Ejtehadi et al., 2004). Other proteins such as FKBP and CI2 had larger barriers of 25 kJ/mol and 42 kJ/mol, respectively (Ejtehadi et al., 2004). The large barriers observed here also suggest that many-body interactions may be playing a significant role in the energy function. A variational theory for the free energy surface of λ-repressor gave a barrier of $\sim$12 kJ/mol (Portman et al., 2001). All-atom simulations of a three-helix bundle fragment of protein A in explicit water gave barrier heights $\approx$17 kJ/mol at the transition midpoint (Garcia and Onuchic, 2003). Applying Kramers theory with an experimentally determined estimate for the prefactor gave an estimate for the free energy barrier of $\sim$18 kJ/mol for the cold shock protein CspTm (Schuler et al., 2002). An analysis which took prefactors from experimental data, along with a thermodynamic analysis to extract enthalpic and entropic contributions to the barrier, gave typical barrier heights of $\sim$30 kJ/mol for the proteins analyzed (Akmal and Munoz, 2004). However, these last two methods found barrier heights under conditions of zero denaturant—the barrier heights at zero stability would likely be significantly higher. For example, the average $\langle (m_{U\ddagger}/m)\Delta G \rangle$ for the proteins in Table 1 is $\sim$17 kJ/mol, to be added to the barrier height at conditions of zero denaturant.

Applying this method to a simulation model, where one knows the answers in advance, provides a good control for the study and is a topic for future work.

REFERENCES

Akmal, A., and V. Munoz. 2004. The nature of the free energy barriers to two state folding. *Proteins*. 57:142–152.

Bryngelson, J. D., J. N. Onuchic, N. D. Socci, and P. G. Wolynes. 1995. Funnels, pathways and the energy landscape of protein folding. *Proteins*. 21:167–195.

Bryngelson, J. D., and P. G. Wolynes. 1989. Intermediates and barrier crossing in a random energy model (with applications to protein folding). *J. Phys. Chem.* 93:6902–6915.

CRC. 2003. CRC Handbook of Chemistry and Physics, 84th Ed. D.R. Lide, editor. CRC Press, New York.

D'Aquino, J. A., J. Gomez, V. J. Hilser, K. H. Lee, L. M. Amzel, and E. Freire. 1996. The magnitude of the backbone conformational entropy change in protein folding. *Proteins Struct. Funct. Gen.* 25:143–156.

Eaton, W. A., V. Munoz, S. J. Hagen, G. S. Jas, L. J. Lapidus, E. R. Henry, and J. Hofrichter. 2000. Fast kinetics and mechanisms in protein folding. *Annu. Rev. Biophys. Biomol. Struct.* 29:327–359.

Ejtehadi, M. R., S. P. Avall, and S. S. Plotkin. 2004. Three-body interactions improve the prediction of rate and mechanism in protein folding models. *Proc. Natl. Acad. Sci. USA.* 101:15088–15093.

Fersht, A. R. 1999. Structure and Mechanism in Protein Science, 1st Ed. W.H. Freeman, New York.

Garcia, A. E., and J. N. Onuchic. 2003. Folding a protein in a computer: an atomic description of the folding/unfolding of protein A. *Proc. Natl. Acad. Sci. USA.* 100:13898–13903.

Hanggi, P., P. Talkner, and M. Borkevec. 1990. Reaction-rate theory: fifty years after Kramers. *Rev. Mod. Phys.* 62:251–341.

Jackson, S. E., and A. R. Fersht. 1991. Folding of chymotrypsin inhibitor 2. 1. Evidence for a two state transition. *Biochemistry.* 30:10428–10435.

Klimov, D. K., and D. Thirumalai. 1997. Viscosity dependence of the folding rates of proteins. *Phys. Rev. Lett.* 79:317–320.

Kuhlman, B., D. L. Luisi, P. A. Evans, and D. P. Raleigh. 1998. Global analysis of the effects of temperature and denaturant on the folding and unfolding kinetics of the N-terminal domain of the protein l9. *J. Mol. Biol.* 284:1661–1670.

Lapidus, L. J., W. A. Eaton, and J. Hofrichter. 2000. Measuring the rate of intramolecular contact formation in polypeptides. *Proc. Natl. Acad. Sci. USA.* 97:7220–7225.

Leach, S. J., G. Nemethy, and H. A. Scheraga. 1966. Computation of sterically allowed conformations of peptides. *Biopolymers.* 4:369–407.

Li, M. S., D. K. Klimov, and D. Thirumalai. 2004. Thermal denaturation and folding rates of single domain proteins: size matters. *Polym.* 45:573–579.

Lindberg, M., J. Tangrot, and M. Oliveberg. 2002. Complete change of the protein folding transition state upon circular permutation. *Nat. Struct. Biol.* 9:818–822.

Mines, G. A., T. Pascher, S. C. L. J. R. Winkler, and H. B. Gray. 1996. Cytochrome-c folding triggered by electron transfer. *Chem. Biol.* 3:491–497.

Onuchic, J. N., Z. Luthey-Schulten, and P. G. Wolynes. 1997. Theory of protein folding: the energy landscape perspective. *Annu. Rev. Phys. Chem.* 48:545–600.

Otzen, D. E., and M. Oliveberg. 2004. Correspondence between anomalous m- and δc_p-values in protein folding. *Protein Sci.* 13:3253–3263.

Perl, D., M. Jacob, M. Bano, M. Stupak, M. Antalik, and F. X. Schmid. 2002. Thermodynamics of a diffusional protein folding reaction. *Biophys. Chem.* 96:173–190.

Plotkin, S. S., and J. N. Onuchic. 2000. Investigation of routes and funnels in protein folding by free energy functional methods. *Proc. Natl. Acad. Sci. USA.* 97:6509–6514.

Plotkin, S. S., and J. N. Onuchic. 2002a. Understanding protein folding with energy landscape theory. I. Basic concepts. *Q. Rev. Biophys.* 35:111–167.

Plotkin, S. S., and J. N. Onuchic. 2002b. Understanding protein folding with energy landscape theory. II. Quantitative aspects. *Q. Rev. Biophys.* 35:205–286.

Portman, J. J., S. Takada, and P. G. Wolynes. 2001. Microscopic theory of protein folding rates. II. Local reaction coordinates and chain dynamics. *J. Chem. Phys.* 114:5082–5096.

Scalley, M., and D. Baker. 1997. Protein folding kinetics exhibit an Arrhenius temperature dependence when corrected for the temperature dependence of protein stability. *Proc. Natl. Acad. Sci. USA.* 94:10636–10640.

Schindler, T., and F. X. Schmid. 1996. Thermodynamic properties of an extremely rapid protein folding reaction. *Biochemistry.* 35:16833–16842.

Schuler, B., E. A. Lipman, and W. A. Eaton. 2002. Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature.* 419:743–747.

Shea, J. E., and C. L. Brooks III. 2001. From folding theories to folding proteins: a review and assessment of simulation studies of protein folding and unfolding. *Annu. Rev. Phys. Chem.* 52:499–535.

Shoemaker, B. A., J. Wang, and P. G. Wolynes. 1999. Exploring structures in protein folding funnels with free energy functionals: the transition state ensemble. *J. Mol. Biol.* 287:675–694.

Socci, N. D., J. N. Onuchic, and P. G. Wolynes. 1996. Diffusive dynamics of the reaction coordinate for protein folding funnels. *J. Chem. Phys.* 104:5860–5868.

Wang, J., S. S. Plotkin, and P. G. Wolynes. 1997. Configurational diffusion on a locally connected correlated energy landscape; application to finite, random heteropolymers. *J. Phys. I (Fr.).* 7:395–421.

Zhou, Y., C. Zhang, G. Stell, and J. Wang. 2003. Temperature dependence of the distribution of the first passage time: results from discontinuous molecular dynamics simulations of an all-atom model of the second β-hairpin fragment of protein-G. *J. Am. Chem. Soc.* 125:6300–6305.