Determination of barrier heights and prefactors from protein folding rate data

S. S. Plotkin[‡*]

‡ *Department of Physics and Astronomy,*
University of British Columbia,
Vancouver, BC V6T-1Z1, Canada

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.

I. 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 vs. 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):

$$\Delta G_{\text{F}\ddagger}(T,c) \equiv G_{\ddagger}(T,c) - G_{\text{F}}(T,c) = \Delta G_{\text{F}\ddagger}(T,0) - m_{\text{F}\ddagger}c \tag{I1a}$$

$$\Delta G_{\text{U}\ddagger}(T,c) \equiv G_{\ddagger}(T,c) - G_{\text{U}}(T,c) = \Delta G_{\text{U}\ddagger}(T,0) + m_{\text{U}\ddagger}c \tag{I1b}$$

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

Subtracting I1b from I1a, and defining $\Delta G \equiv \Delta G_{\text{FU}} = G_{\text{U}} - G_{\text{F}}$ and $m = m_{\text{F}\ddagger} + m_{\text{U}\ddagger}$ we have that

$$\Delta G(T,c) = \Delta G(T,0) - mc \tag{I2}$$

For 2-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

$$\begin{aligned} \ln k_{\text{F}}(T,c) &= \ln k_o(T,c) - \Delta G_{\text{U}\ddagger}(T,c)/T \\ &= \ln k_o(T,c) - (\Delta G_{\text{U}\ddagger}(T,0) + m_{\text{U}\ddagger}c)/T \end{aligned} \tag{I3}$$

Eliminating c from equations I2 and I3 gives

$$\ln k_{\text{F}}(T,c) - \frac{m_{\text{U}\ddagger}}{m}\frac{\Delta G(T,c)}{T} = \ln k_o(T,c) - \frac{1}{T}\left(\Delta G_{\text{U}\ddagger}(T,0) + \frac{m_{\text{U}\ddagger}}{m}\Delta G(T,0)\right), \tag{I4}$$

where the left hand side of I4 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 et al (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 vs. $1/T$.

*Electronic address: steve@physics.ubc.ca

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 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:

$$\delta \ln k_{\text{F}} \equiv \ln k_{\text{F}}(T,c) - \ln k_{\text{F}}(T,0) = -\left(\Delta G_{\text{U}\ddagger}(T,c) - \Delta G_{\text{U}\ddagger}(T,0)\right)/T = -\delta \Delta G_{\text{U}\ddagger}/T \tag{I5}$$

which together with eq. I4 gives

$$\delta \Delta G_{\text{U}\ddagger} = -(m_{\text{U}\ddagger}/m)\delta \Delta G \tag{I6}$$
$$\delta \ln k_{\text{F}} = (m_{\text{U}\ddagger}/m)(\delta \Delta G/T) . \tag{I7}$$

This quantifies the assertion above that log folding rates depend linearly on the relative stability of the products. If we let $m_{\text{U}\ddagger}/m \equiv Q^{\neq}$, eq. I6 can be rewritten as

$$\delta G_{\ddagger} = Q^{\neq}\delta G_{\text{F}} + (1 - Q^{\neq})\delta G_{\text{U}} \tag{I8}$$

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

Inspection of rate-stability isotherms for several different proteins (cytochrome C (Mines et al., 1996), protein L (Scalley and Baker, 1997), cspB (Schindler and Schmid, 1996), N-terminal protein L9 (Kuhlman et al., 1997), S6 (Otzen and Oliveberg, 2004)) shows linearity over ranges up to $\approx 25\text{kJ} \cdot \text{mol}^{-1} \approx 10k_{\text{B}}T$, indicating large and robust folding barriers- substantially larger than the folding barriers seen in many simulations for example (c.f. figure 1).

At a higher temperature, the log rate vs. 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 2 effects: 1.) the prefactor increases at higher temperature (since activated escape from traps is further facilitated, and solvent viscosity is reduced), and 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.

II. 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):

$$\Delta G(T,c) = \Delta H - T\Delta S + \Delta C_{\text{P}}\left(T - T_o - T\ln\left(T/T_o\right)\right) . \tag{II9}$$

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

$$\Delta C_{\text{P}}\left(T - T_o - T\ln\left(T/T_o\right)\right) = \left(T - T_o\right)\Delta S + m\left(c - c_o\right) . \tag{II10}$$

For two-state folders, the heat capacity ratio $\Delta C_{\text{P U}\ddagger}/\Delta C_{\text{P}}$ is approximately equal to m-value ratio $-m_{\text{U}\ddagger}/m$, giving the fractional solvent accessibility of the transition state. We assume this equality here as well, which gives for eq. II10:

$$\Delta C_{\text{P U}\ddagger}\left(T - T_o - T\ln\left(T/T_o\right)\right) = -\frac{m_{\text{U}\ddagger}}{m}\Delta S\left(T - T_o\right) - m_{\text{U}\ddagger}\left(c - c_o\right) . \tag{II11}$$

Inserting eq. II11 into Gibbs-Helmholtz expressions for the barrier heights $\Delta G_{\text{U}\ddagger}$ at (T,c) and (T_o, c_o) gives the change in barrier height *at fixed stability*:

$$\left[\Delta G_{\text{U}\ddagger}(T,c) - \Delta G_{\text{U}\ddagger}(T_o,c_o)\right]_{\Delta G(T,c)=\Delta G(T_o,c_o)} \equiv \delta'\Delta G_{\text{U}\ddagger} = -\left(T - T_o\right)\left(\Delta S_{\text{U}\ddagger} + \frac{m_{\text{U}\ddagger}}{m}\Delta S\right) , \tag{II12}$$

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 figure 1 for example.

For changes in temperature of a few degrees, the change in barrier height $\delta'\Delta G_{\text{U}\ddagger}$ is only a few percent of the total barrier height, when the rates vs. temperature and denaturant are fit to a model to extract thermodynamic

parameters, as in ref.s (Kuhlman et al., 1997: Otzen and Oliveberg, 2004: Scalley and Baker, 1997: Schindler and Schmid, 1996). We used equation II12 for the change in barrier height when thermodynamic data were available. For the case of cytC we set $\delta'\Delta G = 0$.

The rates for pairs of states at the same stability ΔG are given from eq. I3 as

$$\ln k_{\mathrm{F}}(T_o, c_o) = \ln k_o(T_o) - \Delta G_{\mathrm{U\ddagger}}(T_o, c_o)/T_o \tag{II13a}$$

$$\ln k_{\mathrm{F}}(T, c) = \ln k_o(T) - \Delta G_{\mathrm{U\ddagger}}(T_o, c_o)/T - \delta'\Delta G_{\mathrm{U\ddagger}}/T \tag{II13b}$$

A. 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 equation I3 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

$$\ln k_o(T_o) = \ln k_{oo} - \Delta^2/2T_o^2 \tag{II14a}$$

$$\ln k_o(T) = \ln k_{oo} - \Delta^2/2T^2 + \ln\left(\eta(T_o)/\eta(T)\right) . \tag{II14b}$$

To compare rate theories with experimental data we must introduce a fundamental time scale or rate constant k_{oo}, which is then modified by barriers representing the ruggedness of the energy landscape. Rates for short loop closure are about $2 \times 10^7 \mathrm{s}^{-1}$ (Lapidus et al., 2000), comparable to helix formation rates of $\sim 10^7 \mathrm{s}^{-1}$, and somewhat faster than rates of hairpin formation $\sim 10^6 \mathrm{s}^{-1}$ (Eaton et al., 2000). We take $10^7 \mathrm{s}^{-1}$ 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 $k_{oo} = 10^9 \mathrm{s}^{-1}$. We will see later that larger estimates for k_{oo} give larger 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 II14a and II14b to give an equation relating the prefactors:

$$\ln k_o(T) = \left(1 - \frac{T_o^2}{T^2}\right)\ln k_{oo} + \frac{T_o^2}{T^2}\ln k_o(T_o) + \ln\left(\frac{\eta(T_o)}{\eta(T)}\right) . \tag{II15}$$

Equations II13a, II13b, and II15 constitute a system of 3 linear equations for 3 unknowns: $\Delta G_{\mathrm{U\ddagger}}(T_o, c_o)$, $\ln k_o(T_o)$ and $\ln k_o(T)$, which can be solved analytically at any given stability, from linear fits to the log rate-stability data.

III. RESULTS

The results of applying the method are shown in figure 2, for the data in figure 1, ranging from the stability of wild type at $296K$ (-74kJ/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. I6 held until the transition midpoint, the barrier would be about 30 kJ/mol plus whatever the barrier was at conditions of zero denaturant.

The slope $\delta\Delta G_{\mathrm{U\ddagger}}/\delta\Delta G \approx 0.8$ is also larger than its empirical value of $m_{\mathrm{U\ddagger}}/m \approx 0.4$ (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 2 parameters in the theory for which we have put in approximate values: the value of the attempt frequency $k_{oo} = 10^9 \mathrm{s}^{-1}$, and the value of $\delta'\Delta G_{\mathrm{U\ddagger}}$, which we have set to zero for cytochrome C in the absence of an empirically determined value. Increasing k_{oo} or decreasing $\delta'\Delta G_{\mathrm{U\ddagger}}$ raises barriers, but does not change the slope $\delta\Delta G_{\mathrm{U\ddagger}}/\delta\Delta G$. The value of $-\Delta G$ where the barrier vanishes linearly decreases as $\delta'\Delta G_{\mathrm{U\ddagger}}$ is decreased below zero, with the barrier vanishing at the stability of the wild type when $\delta'\Delta G_{\mathrm{U\ddagger}}$ is about -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 $\delta'\Delta G_{\mathrm{U\ddagger}}$. We will see later that this sensitivity is not present when a correlated landscape model is used for the prefactor.

Figure 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 vs. temperature.

Equation II14a or II14b may now be solved for Δ, giving a number ≈ 15 kJ/mol, that only weakly depends on stability ΔG or barrier change $\delta'\Delta G_{\mathrm{U}\ddagger}$. Estimating the chain conformational entropy as $\sim 100k_{\mathrm{B}}$ (D'Aquino et al., 1996: Leach et al., 1966), we can give an estimate for the glass temperature T_{G} for this system,

$$T_{\mathrm{G}} = \Delta/(2S_o/k_{\mathrm{B}})^{1/2} \tag{III16}$$

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

A. 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

$$\ln k_o(T_o) = \ln k_{oo} - (S^{\neq}/2)\left(\alpha - \beta\left(1 - T_{\mathrm{G}}/T_o\right)^2\right) \tag{III17a}$$

$$\ln k_o(T) = \ln k_{oo} - (S^{\neq}/2)\left(\alpha - \beta\left(1 - T_{\mathrm{G}}/T\right)^2\right) + \ln\left(\eta(T_o)/\eta(T)\right) . \tag{III17b}$$

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 vs. stability, $m_{\mathrm{U}\ddagger}/m$. In table I this number is compared to the value of T_{G} that emerges from the REM analysis. A mismatch of these 2 values may indicate a breakdown of the REM approximation for states in determining prefactors, i.e. a breakdown in the validity of eq.s II14a,b. For cyt C the value of T_{G} giving the correct slope is about 1.2 kJ/mol, vs. 1.0 kJ/mol from the REM analysis.

The entropy may be eliminated from III17a and III17b, giving an equation that relates the prefactors, and replacing eq. II15:

$$\left[\ln k_{oo} - \ln k_o(T) + \ln\left(\eta(T_o)/\eta(T)\right)\right]\left[\alpha - \beta\left(1 - T_{\mathrm{G}}/T_o\right)^2\right] = \left[\ln k_{oo} - \ln k_o(T_o)\right]\left[\alpha - \beta\left(1 - T_{\mathrm{G}}/T\right)^2\right] \tag{III18}$$

Equations II13a, II13b, and III18 again define a system of 3 linear equations for 3 unknowns: $\Delta G_{\mathrm{U}\ddagger}(T_o, c_o)$, $\ln k_o(T_o)$ and $\ln k_o(T)$, which may be solved analytically. The results are shown in figure 4.

We see that both barriers and prefactors are larger than the corresponding REM values, and the analysis for other proteins yields quite large numbers in general (c.f. table I for numbers). The barriers at the transition midpoint are about $22k_{\mathrm{B}}T_{300K}$, and prefactors are almost unactivated. The REM value of T_{G} resulted from approximating a value of $100k_{\mathrm{B}}$ 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 $m_{\mathrm{U}\ddagger}/m$. The parameters α and β could in principle have been adjusted to best match the experimental slope, however it can be shown that this results in the same solution of II13a, II13b, and III18 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 (figure 4). Nor were there any significant effects due to barrier height difference- as $\delta'\Delta G_{\mathrm{U}\ddagger}$ changed from -2 kJ/mol to 0 kJ/mol, the barrier changed by less than 2%. The effects due to T_{G} are modest as well: over the range of T_{G} values in figure 3B, the barrier height changed by less than 15%. Lastly, the prefactors of the correlated landscape model are nearly constant over the range of experimental stabilities (figure 4), consistent with empirical observations (c.f. the comments below eq. I4).

Equation III17a or III17b may now be solved for $S^{\neq}$ as a check, giving $S^{\neq} \approx 40k_{\mathrm{B}}$, or about 40% 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 $S^{\neq} \approx (1 - m_{\mathrm{U}\ddagger}/m)S_o$, then eq. III16 gives $\Delta \approx 14$ kJ/mol. Since the variances of individual residues add to give Δ^2, $\Delta^2 \approx N(1 - m_{\mathrm{U}\ddagger}/m)b^2$, where b is a non-native energy scale per residue, here $\approx 0.7k_{\mathrm{B}}T_{300}$.

Figure 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 roughly equal. The barriers at the transition midpoint $\Delta G_{\mathrm{U}\ddagger}^o$ and at the stability of the wild-type protein $\Delta G_{\mathrm{U}\ddagger}^{(wt)}$ increase linearly, as $\sim 2T_o \ln k_{oo}$.

In the REM analysis there is an intermediate regime where the prefactor has a more complex temperature dependence than eq. II14a. We do not describe this regime in detail since it is obtained from eq.s III17a and III17b 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 about 11 kJ/mol. Similarly the prefactor monotonically increases from $10^8\mathrm{s}^{-1}$ at the stability of the wild-type to $10^{10}\mathrm{s}^{-1}$ at zero stability. We note that these problems are not present if the stability difference $\delta'\Delta G_{\mathrm{U\ddagger}}$ is set to zero, if the prefactor is 2 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., 1997: Otzen and Oliveberg, 2004: Scalley and Baker, 1997: Schindler and Schmid, 1996) is likely to give more accurate numbers. Likewise in the correlated model for NTL9, the prefactor increased from about $10^8\mathrm{s}^{-1}$ at the stability of the wild-type to unphysical values at zero stability. A similar situation exists in the REM 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 vs. 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. 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 (about $10^2\mathrm{s}^{-1}$). Such small barriers are consistent with estimates taken from simulations using C_α-models (Shea and Brooks III, 2001).

B.   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 eq.s II14a and II14b are replaced by

$$\ln k_o(T_o) = \ln k_{oo} - E_{\mathrm{A}}/T_o \tag{III19a}$$
$$\ln k_o(T) = \ln k_{oo} - E_{\mathrm{A}}/T + \ln\left(\eta(T_o)/\eta(T)\right)\;, \tag{III19b}$$

from which E_{A} may be eliminated yielding

$$\ln k_o(T) = (1 - T_o/T)\ln k_{oo} + (T_o/T)\ln k_o(T_o) + \ln\left(\eta(T_o)/\eta(T)\right)\;. \tag{III20}$$

This equation relating the prefactors together with eq.s II13a and II13b are the new system of equations to be solved.
Eliminating $\Delta G_{\mathrm{U\ddagger}}$ from II13a and II13b gives another equation relating the prefactors:

$$\ln k_o(T) = \ln k_{\mathrm{F}}(T,c) - (T/T_o)\ln k_{\mathrm{F}}(T_o,c_o) + \delta'\Delta G_{\mathrm{U\ddagger}}/T + (T_o/T)\ln k_o(T_o)\;. \tag{III21}$$

Equations III21 and III20 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 then gives an equation that is independent of any variable to be solved for:

$$\ln k_{\mathrm{F}}(T,c) - (T/T_o)\ln k_{\mathrm{F}}(T_o,c_o) + \delta'\Delta G_{\mathrm{U\ddagger}}/T = (1 - T_o/T)\ln k_{oo} + \ln\left(\eta(T_o)/\eta(T)\right) \tag{III22}$$

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 3 equations in 3 variables is equivalent to finding the point where 3 planes intersect. Letting

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

equations II13a, II13b, and III20 may be recast as

$$x_2 - (T_o/T)x_1 = A \tag{III23a}$$
$$x_2 - (T_o/T)x_1 = B \tag{III23b}$$
$$x_1 - (1/T_o)x_3 = C \tag{III23c}$$

where

$$
\begin{aligned}
A &= (1 - T_o/T)\ln k_{oo} + \ln(\eta(T_o)/\eta(T)) \\
B &= \ln k_{\mathrm{F}}(T_o, c_o) + (T_o/T)\ln k_{\mathrm{F}}(T, c) + \delta'\Delta G_{\mathrm{U}\ddagger}/T \\
C &= \ln k_{\mathrm{F}}(T_o, c_o)\,.
\end{aligned}
$$

Since $A \neq B$ in general, eq.s III23a and III23b 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.

IV. 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 3 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 dressing 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 about $19k_{\mathrm{B}}T$, and the corresponding barriers in the correlated model is about $18k_{\mathrm{B}}T$. If CspB is omitted the barriers are $21k_{\mathrm{B}}T$ and $22k_{\mathrm{B}}T$ 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 ($10^9\mathrm{s}^{-1}$). 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 2-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 vs. stability curve had to remain roughly 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 parameters must themselves become temperature-dependent. Nevertheless, the fact that the Arrhenius activation model fails in general to provide a solution for even 2 temperatures (2 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 equations II14a, III17a, or III19a 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 $\Delta H_{\mathrm{U}\ddagger}$ (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., 1997: Scalley and Baker, 1997) by this method. However 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. (II12)), 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...). All rate-stability curves for a given protein must be 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 I on the experimental data.

Other methods have been used to estimate barrier heights. Adding a 3-body contribution to a pair-wise interacting energy function to give best agreement with experimental ϕ-values, a barrier height for protein L of about 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 approximately 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 about 18 kJ/mol for the cold shock protein CspTm (B.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 about 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 $\langle (m_{\mathrm{U}\ddagger}/m)\Delta G\rangle$ for the proteins in table I is about 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.

V.   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.

B.Schuler, 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.

CRC. 2003. *In* CRC Handbook of Chemistry and Physics. D. R. Lide, editor, 84th Ed. 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-structure Function Genetics*. 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 and Biomolec 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 Nat Acad Sci USA*. 101:15088–15093.

Fersht, A. R. 1999. Structure and mechanism in protein science, 1st Ed. W. H. Freeman and Co., 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 Nat 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. 1997. 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 Nat 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.

Lindberg, M., J. Tangrot, and M. Oliveberg. 2002. Complete change of the protein folding transition state upon circular permutation. *nsb*. 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. and 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 Nat Acad Sci USA.* 97:6509–6514.

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

Plotkin, S. S., and J. N. Onuchic. 2002b. Understanding protein folding with energy landscape theory ii: Quantitative aspects. *Quart. 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 Nat 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.

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. *Annual Review of Physical Chemistry.* 52:499–535.

Shoemaker, B. A., J. Wang, and P. G. Wolynes. 1999. Exploring structures in protein folding funnels with free energy functinals: 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 France.* 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 beta-hairpin fragment of protein g. *J Am Chem Soc.* 125:6300–6305.

FIGURE CAPTIONS

FIGURE 1: Logarithm of the rate *vs.* (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.* (Mines et al., 1996).

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 cytochrome C. The wild type protein has a stability of $\Delta G \approx 74$ kJ/mol. Numerical values are given in table I. 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.

FIGURE 3: **(A)** The temperature T_G that emerges from the REM analysis for cyt-C (see text and eq. III16) 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 wildtype (37 kJ/mol). **(B)** T_G also changes little as native stability ΔG is varied (for this plot $\delta' \Delta G_{U\ddagger} = 0$).

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 cytC. Numerical values are given in table I. 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 roughly constant and solvent viscosity plays a minor role.
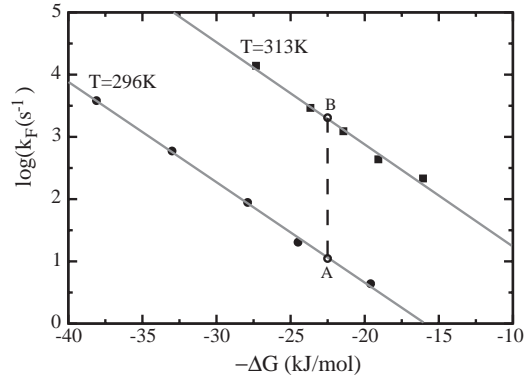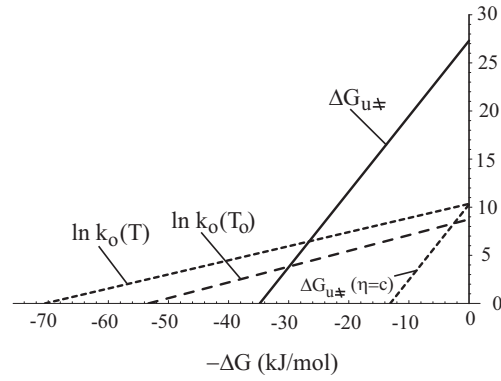
FIG. 1:

FIG. 2:

FIGURE 5: Barrier heights and prefactors extracted from the recipe for the correlated energy landscape (see text) increase as the bare reconfiguration rate (defined in III17a and III17b) increases. The increase is linear. $\Delta G^o_{U\ddagger}$ is the barrier at the transition midpoint, $\Delta G^{(wt)}_{U\ddagger}$ is the barrier at the stability of the wild-type protein, and $k_o(T_o)$ is the prefactor at temperature T_o in s^{-1}.
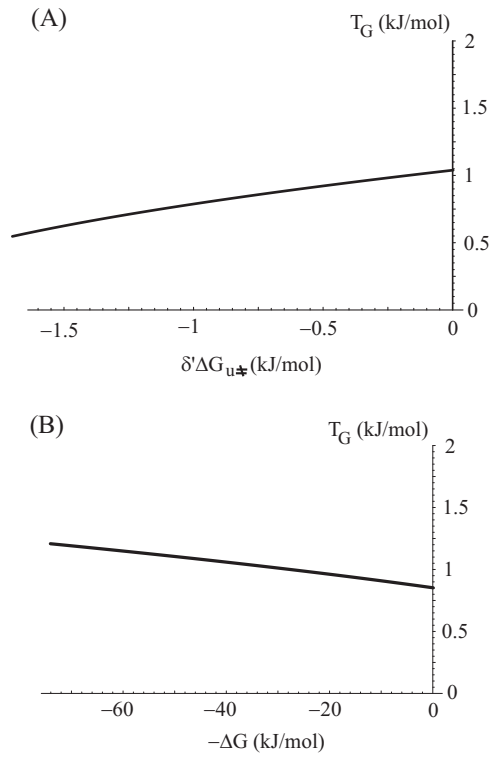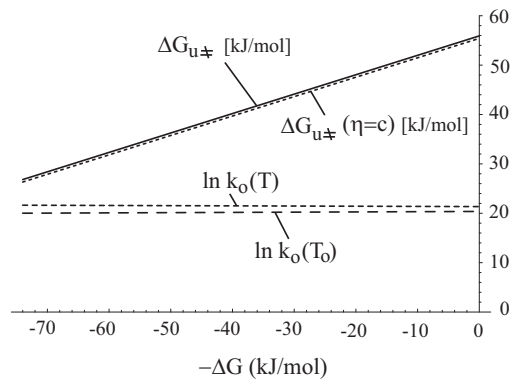
(A)

T_G (kJ/mol)

$\delta'\Delta G_{u\ddagger}$ (kJ/mol)

(B)

T_G (kJ/mol)

$-\Delta G$ (kJ/mol)

FIG. 3:

$\Delta G_{u\ddagger}$ [kJ/mol]

$\Delta G_{u\ddagger}$ (η=c) [kJ/mol]

$\ln k_o(T)$

$\ln k_o(T_o)$

$-\Delta G$ (kJ/mol)

FIG. 4:

$\Delta G_{u\ddagger}^0$ [kJ/mol]

$\Delta G_{u\ddagger}^{(wt)}$ [kJ/mol]

$\ln k_o(T_o)$

$\ln (k_{00}$ (s^{-1}))
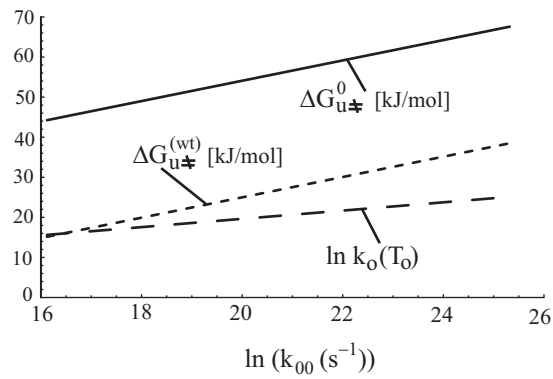
FIG. 5:

TABLE I: **Thermodynamic and kinetic parameters for proteins studied**

Proteins [a]	T_o	T	ΔG^{WT} [b]	$\delta'\Delta G_{\mathrm{U}\ddagger}$	Correlated Model				Random Energy Model			
					$\Delta G^o_{\mathrm{U}\ddagger}$ [c]	$\Delta G^{wt}_{\mathrm{U}\ddagger}$ [d]	$k_o(T_o)$ [c]	$T_{\mathrm{G}}^{(m)}$ [e]	$\Delta G^o_{\mathrm{U}\ddagger}$ [c]	$\Delta G^{wt}_{\mathrm{U}\ddagger}$ [d]	$k_o(T_o)$ [c]	$T_{\mathrm{G}}^{\mathrm{REM}}$ [f]
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) [g]
S6	2.48	2.56	-31	-1.4	61	39	9×10^8	1.2	78	21	(10^{12})	0 (0.7) [g]
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 [g]	24	20	2×10^5	0.7

[a]Sources for experimental data: cyt C (Mines et al., 1996), NTL9 (Kuhlman et al., 1997), S6 (Otzen and Oliveberg, 2004), PTL (Scalley and Baker, 1997), cspB (Schindler and Schmid, 1996). All temperatures and energies are in kJ/mol. All rates are in s^{-1}.

[b]Stability of the wild type protein.

[c]At the transition midpoint where $\Delta G = 0$.

[d]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.

[e]Value of T_{G} that gives a slope of barrier height vs. stability equivalent to the experimental value of $m_{\mathrm{U}\ddagger}/m$.

[f]Value of T_{G} using the REM approximation for rates, taken at a stability of about 1/2 of the wild-type protein.

[g]See text for explanation and comments.