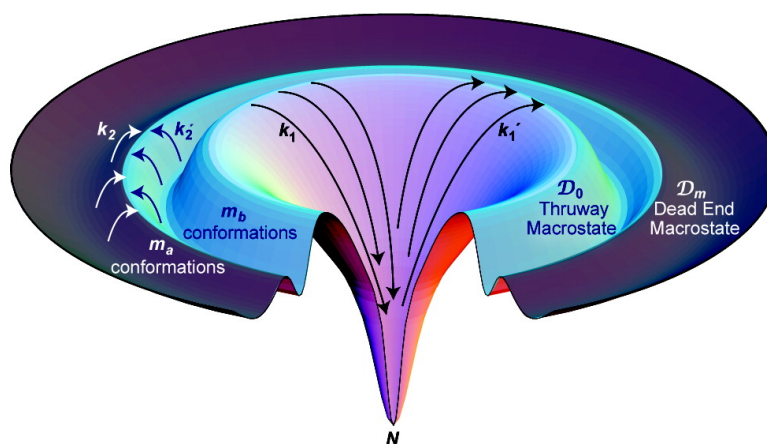Article

# The Ultimate Speed Limit to Protein Folding Is Conformational Searching

Kingshuk Ghosh, S. Banu Ozkan, and Ken A. Dill

## More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 6 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

# The Ultimate Speed Limit to Protein Folding Is Conformational Searching

Kingshuk Ghosh, S. Banu Ozkan,† and Ken A. Dill*

*Contribution from the Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94158*

Received September 20, 2006; E-mail: dill@maxwell.ucsf.edu

**Abstract:** More than a dozen proteins are known to be ultrafast folders. In addition to being fast, their kinetics is unusual. Like traditional rate processes, fast folding proteins have activation barriers at low temperatures, but unlike traditional processes, they have negative activation energies at high temperatures. We develop a model of ultrafast folders that joins a macroscopic mass-action model with a microscopic energy landscape description; we call it the *Thruway Search Model*. We find good agreement with experimental rates and equilibria on 13 ultrafast folders. The observed folding rates are found to be proportional to the number of microscopic folding routes: fast-folding proteins have more parallel microscopic routes on energy landscapes. At high temperatures, where traditional barriers are small, the remaining bottleneck is a search through denatured conformations to find thruway routes to the native state. Negative activation arises because increasing temperature expands the denatured ensemble, broadening the search, slowing the folding to the native state. We find that the upper estimate of the free energy barriers are positive but small, as little as 0.5 kT.

## I. Introduction

Some proteins fold very rapidly. Called *ultrafast* folders, more than a dozen proteins known so far fold up on the microsecond time scale.[1–16] This time scale is interesting because it is regarded as the speed limit of folding:[17] it is the fastest speed that a protein can fold when it has no thermal barriers. Ultrafast folders can give insights into the intrinsic physical limitations on folding speeds.

The thermal and kinetic properties of ultrafast folders are interesting and puzzling. Most ultrafast folders show single-exponential kinetics,[1,5,7–11,13,16] but some fold with more complex kinetics.[3,4,18] Equally puzzling, Arrhenius plots of their folding rates vs temperature show normal positive activation energy barriers at low temperatures, but they crossover to become *negative activation enthalpies* at high temperatures.[10] Non-Arrhenius kinetics (i.e., a nonlinear dependence of folding rate on temperature) has also been observed for some two-state proteins.[19–28] This non-Arrhenius behavior has been explained in different ways: (i) as a nonlinear temperature dependence of the configurational diffusion constant (i.e., the front factor in the rate equation) on rough energy landscapes,[29] or (ii) as a result of the temperature dependence of the hydrophobic interaction, leading to a heat capacity of activation.[30,31] An

† Present address: Physics Department and Center for Biological Physics, Arizona State University, Tempe, AZ 85287.

(1) Qui, L.; Pabitt, S.; Roitberg, A.; Hagen, S. *J. Am. Chem. Soc.* **2002**, *124*, 12952−12953.
(2) Wittung-Stafshede, P.; Lee, J. C.; Winkler, J. R.; Gray, H. B. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 6587−6590.
(3) Snow, C. D.; Nguyen, N.; Pande, V. S.; Gruebele, M. *Nature* **2002**, *420*, 102−106.
(4) Kubelka, J.; Eaton, W. A.; Hofrichter, J. *J. Mol. Biol.* **2003**, *329*, 625−630.
(5) Jager, M.; Nguyen, H.; Crane, J. C.; Kelly, J. W.; Gruebele, M. *J. Mol. Biol.* **2001**, *311*, 373−393.
(6) Spector, S; Raleigh, D. P. *J. Mol. Biol.* **1999**, *293*, 763−768.
(7) Wang, T; Zhu, Y and Gai, F. *J. Phys. Chem. B* **2004**, *108*, 3694−3697.
(8) Yang, W. Y.; Gruebele, M. *Nature* **2003**, *423*, 193−197.
(9) Mayor, U.; Johnson, C. M.; Daggett, V.; Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 13518−13522.
(10) Zhu, Y.; Alonso, D. O. V.; Maki, K.; Huang, C. Y.; Lahr, S. J.; Daggett, V.; Roder, H.; Degrado, W. F.; Gai, F. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 15486−15491.
(11) Myers, J. K.; Oas, T. G. *Nat. Struc. Biol.* **2001**, *8*, 552−558.
(12) Dimitriadis, G.; Drysdale, A.; Myers, J. K.; Arora, P; Radford, S. E.; Oas, T. G.; Smith, D. A. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*(11), 3809−3814.
(13) Nguyen, H.; Jager, M.; Moretto, A.; Gruebele, M.; Kelly, J. W. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*(7), 3948−3953.
(14) Bunagan, M. R.; Yang, X.; Saven, J. G.; Gai, F. *J. Phys. Chem. B* **2006**, *110*, 3759−3763
(15) Yang, W. Y.; Gruebele, M. *Biochemistry* **2004**, *43*, 13018−13025
(16) Xu, Y.; Purkayastha, P.; Gai, F. *J. Am. Chem. Soc.* **2006**, *128*, 15816−15842.
(17) Kubelka, J.; Hofrichter, J.; Eaton, W. A. *Curr. Opin. Struct. Biol.* **2004**, *14*(1), 76−88.

(18) Ma, H.; Gruebele, M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 2283−2287.
(19) Scalley, M. L;Baker, D. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *95*, 10636−10640.
(20) Alexander, P.; Orban, J.; Bryan, P. *Biochemistry* **1992**, *31*, 7243−7248.
(21) Tan, Y. J; Oliveberg, M.; Fersht, A. R. **1996**, *264*, 377−389.
(22) Perl, D.; Jacob, M.; Bano, M.; Stupak. M.; Antalik, M.; Schmid, F. X. *Biophys. Chem.* **2002**, *96*, 173−190.
(23) Nico, A. J.; Van Nuland, J.; Meijbeg, W.; Warner, J.; Forge, V.; Scheek, R. M.; Robillard, G. G.; Dobson, C. M. *Biochemistry* **1998**, *37*, 622−637.
(24) Plaxco, K. W.; Guijarro, J. I.; Morton, C.; Pitkeathly, M.; Campbell, I. D.; Dobson, C. M. *Biochemistry* **1998**, *37*, 2529−2537.
(25) Main, E. R. G.; Fulton, K. F.; Jackson, S. E. *J. Mol. Biol.* **1999**, *291*, 429−444.
(26) Ibarra-Molero, B.; Makhatade, G. I.; Matthews, C. R. *Biochemistry* **2001**, *40*, 719−731.
(27) Manyusa, S.; Whitford, D. *Biochemistry* **1999**, *38*, 9533−9540.
(28) Kuhlman, B.; Luisi, D. L.; Evans, P. A.; Raleigh, D. P. *J. Mol. Biol.* **1998**, *284*(5), 1661−1670.
(29) Bryngelson, J. D.; Onuchic, J. N.; Socci, N. F.; Wolynes, P. G. *Proteins* **1995**, *21*, 167−195.
(30) Chan, H. S; Dill, K. A. *Protein* **1998**, *30*, 2−33.
(31) Akmal, A.; Munoz, V. *Protein* **2004**, *57*(1), 142−52.

interesting observation has been that non-Arrhenius behavior becomes Arrhenius if folding rates are compared under isostability conditions.[19,32,33] However, most of the ultrafast folders,[5,6,8,10,13−15] unlike their fast-folding counterpart (with some exceptions), show significant decreases in their folding rates upon increase in temperature at high temperatures. It means that folding is slowed down, not speeded up, by increasing the temperature. We are interested in the microscopic principles underlying this behavior. It is worth mentioning that the experimentally observed folding rates cited throughout the manuscript are usually extracted from the observed relaxation rates by using a two-state model assumption.

The basis for non-Arrhenius behavior is related to the nature of the landscape.[34] It has been argued that under strong native conditions, the barrier between the folded and unfolded states disappears, and the folding time will be governed by the diffusion rate along the reaction coordinate.[29] Because the chain diffusion is different at different stages of folding, the diffusion coefficient would not be a constant, leading to multiple- or stretched-exponential kinetics, regarded as a signature of downhill folding.[5,8,35] However, barrierless folding does not necessarily imply non-exponential kinetics.[36−40] Downhill folding has been studied using simple protein models.[8,38,39,41] There have also been efforts[8,41] to study relaxation kinetics on 1D energy landscapes based on similar models proposed by Bryngelson et al.[29] using Langevin dynamics simulations.

The model of Munoz and co-workers describes downhill folding as a one-state process at all temperatures. According to that model, the minimum shifts continuously from native to unfolded as the native bias is reduced.[38,42,43] In connection to the study of downhill folding, BBL (the peripheral subunit binding domain from oxoglutarate dehydrogenase) and its three bacterial homologs constitute an interesting system to explore the folding landscape.[42−52]

Here, we develop a model for analyzing experimental temperature-dependent folding kinetics in a way that can give insights into the microscopic energy landscapes that underly them.

## II. Model

On the one hand, we wanted a mass-action-like model that could be used to fit experimental folding and unfolding rate
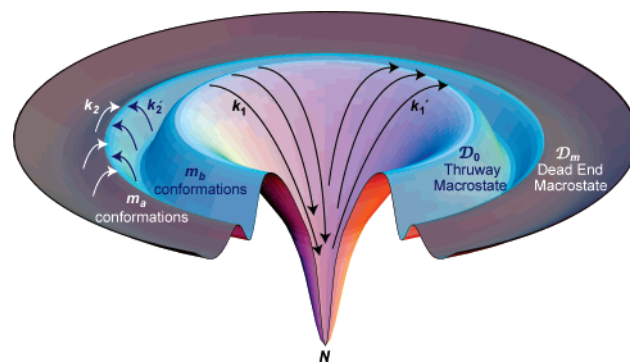


**Figure 1.** Thruway Search Model, microscopic funnel energy landscape. Denatured conformations are either thruway states ($\mathcal{D}_0$ with $m_b$ conformations), having direct access to the native state $N$, or dead-end states ($\mathcal{D}_m$ with $m_a$ conformations), which must search to find thruway routes.

data vs temperature. On the other hand, we also wanted a model that could give insights about features of energy landscapes, such as the numbers of microscopic routes. So, we describe here a hybrid, which we call the *Thruway Search Model*.

We regard the native state $N$ as a single microstate (Figure 1). We divide the denatured macrostate, $\mathcal{D}$, which is a broad ensemble of microstates, into two subsets: (1) Macrostate $\mathcal{D}_0$ (Figure 1) is the set of all the denatured microstates that have direct access to $N$. There are $m_b$ of these conformations, which we call *thruway states*.[53] There are $n$ micro-routes connecting $\mathcal{D}_0$ directly to $N$; we call these *thruway routes*. We do not know $n$ in advance; it is one of the model parameters that we use to fit experimental data.

Macrostate $\mathcal{D}_m$ is the subset of denatured states that are *dead-ends* or kinetic traps; there are $m_a$ such dead-end conformations. To reach the native state, any conformation beginning in $\mathcal{D}_m$ must first pass through a $\mathcal{D}_0$ state. We call these *trap routes*. In classical terminology, the state $\mathcal{D}_m$ would be called an off-pathway kinetic intermediate.

For the process of trapping, in going from any one of the $\mathcal{D}_0$ microstates to any of the $\mathcal{D}_m$ microstates, the rate coefficient is $k_2$. All such micro-routes are assumed identical; we could consider distributions in barrier heights, but it would just lead to unwarranted complexity here. For the reverse process, the rate coefficient is $k_2'$. For the process of folding, in going from any of the $\mathcal{D}_0$ microstates to $N$ along any of the thruway routes, the rate coefficient is $k_1$; the reverse rate coefficient is $k_1'$. $q$ is the number of routes from one $\mathcal{D}_0$ conformation to $N$.

Shown schematically in Figure 1, the master equation for this model is

$$\frac{dD_m}{dt} = -m_b k_2' D_m + m_b k_2 D_0 \tag{1}$$

$$\frac{dD_0}{dt} = -m_a k_2 D_0 + m_a k_2' D_m - q k_1 D_0 \tag{2}$$

$$\frac{dN}{dt} = q k_1 D_0 - q k_1' N \tag{3}$$

The total number of routes $n$ will equal the number of routes $q$, from each thruway state, multiplied by the number of total

(32) Liu, Z.; Chan, H. S. *J. Mol. Biol.* **2005**, *349*, 872−889.
(33) Gutin, A.; Sali, A.; Abkevich, V.; Karplus, M.; Shakhnovich, E. I. *J. Chem. Phys.* **1998**, *108*, 6466−6483.
(34) Zaman, M. H.; Sosnick, T. R.; Berry, S. R. *Phys. Chem. Chem. Phys.* **2003**, *5*(12), 2589−2594.
(35) Sabelko, J.; Ervi, J.; Gruebele, M. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 6031−6036.
(36) Hagen, J. S.; Eaton, W. A. *J. Mol. Biol.* **2000**, *301*, 1019−1021.
(37) Hagen, J. S. *Proteins* **2003**, *50*, 1−4.
(38) Munoz, V. *Int. J. Quantum. Chem.* **2002**, *90*, 1522−1528.
(39) Knott, M; Chan, H. S. *Protein* **2006**, *65*, 373−391.
(40) Hagen, J. S. *Proteins: Struct., Functi., Bioinf.* **2007**, *68*, 205−217.
(41) Ma, H.; Gruebele, M. *J. Comp. Chem.* **2006**, *27*, 125−134.
(42) Olivia, F. Y.; Munoz, V. *J. Am. Chem. Soc.* **2004**, *126*, 8596−8597.
(43) Naganathan, A. N.; Jimenez-Perez, R.; Sanchez-Ruiz, J. M.; Munoz, V. *Biochemistry* **2005**, *44*, 7435−7449.
(44) Garcia-Mira, M. M.; Sadqi, M; Fischer, N; Sanchez-Ruiz, J. M.; Munoz, V. *Science* **2002**, *298*, 2191−2195.
(45) Munoz, V.; Sanchez-Ruiz, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 17646−17651.
(46) Sadqi, M.; Fushman, D.; Munoz, V. *Nature* **2006**, *442*, 317−321.
(47) Sadqi, M.; Fushman, D.; Munoz, V. *Nature* **2007**, *445*, E17.
(48) Kelly, J. W. *Nature* **2006**, *442*, 255−256.
(49) Ferguson, N; Sharpe, T. D.; Schartau, P. J.; Sato, S.; Allen, M. D; Johnson, C. M.; Rutherford, T. J.; Fersht, A. R. *J. Mol. Biol.* **2005**, *353*, 427−446.
(50) Huang, F.; Stao, S.; Shape, T. D.; Ying, L.; Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 123−127.
(51) Ferguson, N.; Sharpe, T. D.; Johnson, C. M.; Schartau, P. J.; Fersht, A. R. *Nature* **2007**, *445*, E14.
(52) Zhou, Z.; Bai, Y. *Nature* **2007**, *445*, E16.
(53) Chan, H. S.; Dill, K. A. *J. Chem. Phys.* **1994**, *100*(12), 9238−9257.
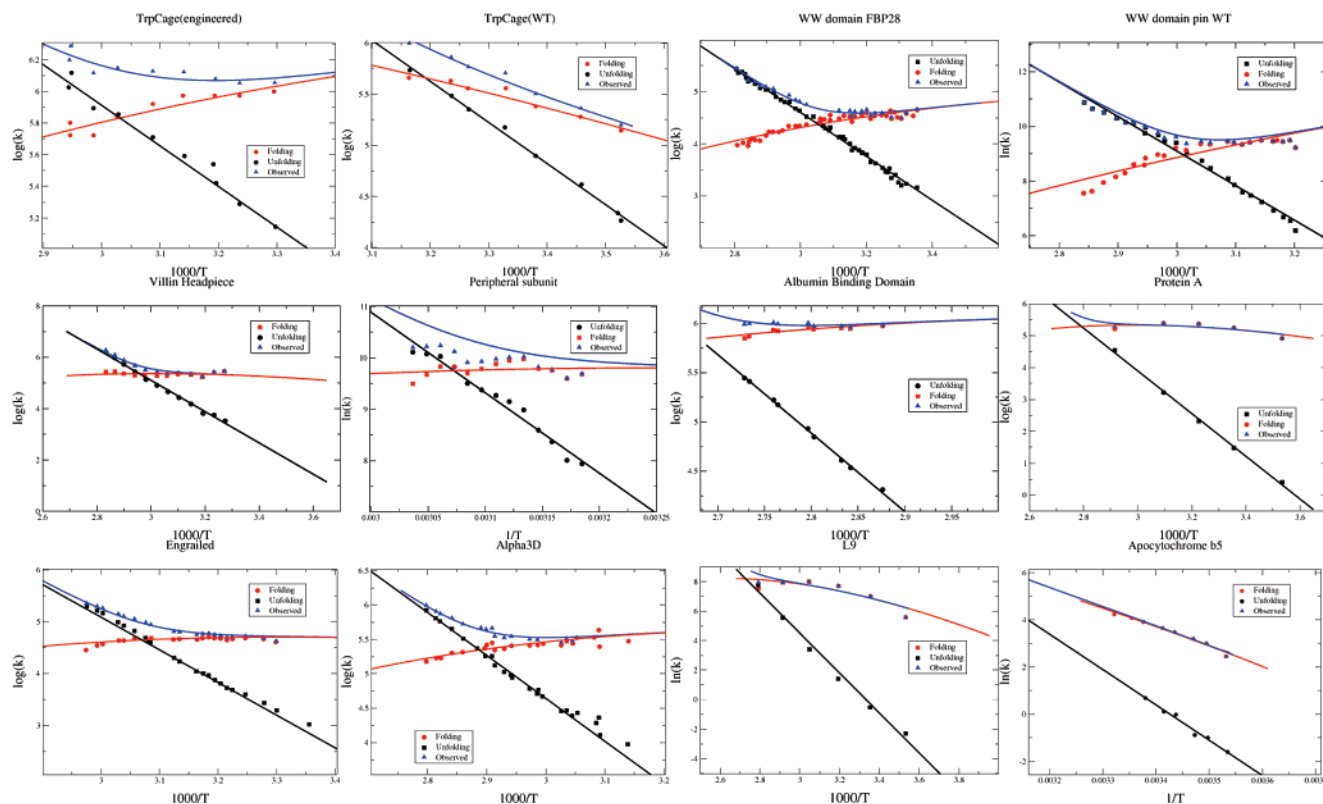
**Figure 2.** Model fits to overall relaxation, folding, and unfolding rate data for different proteins. Overall relaxation or the observed relaxation is in blue, folding rates are in red, and unfolding rates are in black. Solid lines are theoretical fits from the model.

thruway states, $m_b$, $n = qm_b$ (see details in Supporting Information, appendix B). $D_o$, $D_m$ and N in the above equation denote the concentration of macrostates $\mathscr{D}_o$, $\mathscr{D}_m$ and $N$ respectively. A similar set of rate equations has been proposed before,[54] except without the thermal aspects described below. Along each type of route, we assume an Arrhenius barrier,

$$k_1 = k_0 \exp(-\epsilon_{DN}/kT) \qquad (4)$$

$$k_1' = k_0 \exp(-\epsilon_{ND}/kT) \qquad (5)$$

$$k_2 = k_0 \exp(-\epsilon/kT) \qquad (6)$$

$$k_2' = k_0 \exp(-\epsilon/kT) \qquad (7)$$

where $kT$ is Boltzmann's constant multiplied by the absolute temperature and $k_0$ defines an intrinsic time scale, dependent on the chain length, $L$. The energy barrier along the native folding routes from $\mathscr{D}_0$ to N is $\epsilon_{DN}$ and the unfolding barrier is $\epsilon_{ND}$. We assume that the barrier along the trap routes (from $\mathscr{D}_0$ to $\mathscr{D}_m$ or vice versa) is identical in each direction, because this describes a transition from one denatured state to another.

A key aspect here is that the Thruway Search Model treats the expansion of the denatured state with temperature, which leads to increased searching at higher temperatures. We use the Flory−Huggins theory[55,56] to approximate these effects. In our model, the number of denatured conformations, $m(L, T)$, depends on temperature, $T$, and on the chain length, $L$, as described in the appendix A of the Supporting Information.

(54) Ellison, P. A.; Cavagnero, S. *Protein Sci.* **2006**, *15*(3), 564−582.
(55) Dill, K.; Bromberg, S. *Molecular driving forces: statistical thermodynamics in chemistry and biology*; Garland Science: New York, 2003.
(56) Dill, K. A.; Alonso, D. O. V.; Hutchinson, K. *Biochemistry* **1989**, *28*, 5439−5449.

We want to compute the macroscopically observable folding rate, $k_f$, from this microscopic model. So, we calculate the folding time distribution, applying an absorbing boundary condition to the state *N*. The folding rate is the inverse of the average folding time obtained from this distribution. We initiate folding from the fully populated state $\mathscr{D}_0$. (We have also initiated folding from states that are mixtures of $\mathscr{D}_0$ and $\mathscr{D}_m$, but the results are essentially the same, and the present approach keeps the math simpler.) It is shown in the appendix B of the Supporting Information that the folding rate from this model is

$$k_f = \frac{nk_0}{m(L, T)} e^{(-\epsilon_{DN}/kT)} \qquad (8)$$

and the unfolding rate $k_u$ is

$$k_u = nk_0 e^{(-\epsilon_{ND}/kT)} \qquad (9)$$

The unfolding rate is obtained by using $k_1'$, which is the rate coefficient to go from native to $\mathscr{D}_0$ state.

We can also define an overall relaxation rate $k_r$ in the two state picture that can be written as a sum of $k_f$ and $k_u$ defined above. In a typical experiment, it is the overall relaxation rate that is measured directly rather than the folding and unfolding rates separately.

$$k_r = k_f + k_u \qquad (10)$$

Here's how we apply the model. We first compute the intrinsic rate, $k_0$, using the relationship of Eaton et al.,[4] $k_0 = 100/L \, \mu \, \text{s}^{-1}$. Our aim is to fit experimental data for the folding and unfolding rates of a given protein as a function of temperature. To do that, we use the four model parameters: the

***Table 1.*** Extracted Values of the Parameters from the Fit

| protein name | $\epsilon_{DN}(kT)$ | $\epsilon_{ND}(kT)$ | e (kcal/mol) | log($n$) | $L$ |
|---|---|---|---|---|---|
| Trpcage(engineered)[14] | 7.0 | 20.0 | .61 | 6.9 | 20 |
| Trpcage(WT)[1] | 22.1 | 31.4 | 0.74 | 11.8 | 20 |
| WW domain FBP28[13] | 11.7 | 32.9 | 0.51 | 10.7 | 28 |
| WW domain pin(WT)[5] | 10.8 | 42.8 | 0.42 | 13.93 | 34 |
| Villin headpiece subdomain[4] | 24.0 | 47.6 | 0.62 | 16.94 | 35 |
| Peripheral subunit[6] | 25.9 | 53.1 | 0.58 | 18.76 | 41 |
| Albumin binding domain[7] | 28.2 | 62.2 | 0.7 | 20.87 | 47 |
| L9[28] | 29.5 | 45.9 | 1.3 | 13.35 | 56 |
| Protein A[12] | 29.3 | 52.3 | 1.1 | 17.76 | 58 |
| Engrailed Homeo domain[9] | 27.5 | 49.2 | 0.97 | 17.78 | 61 |
| α3D[10] | 26.1 | 48.0 | 1.2 | 16.96 | 73 |
| Apocytochrome b5[27] | 42.5 | 50.8 | 2.1 | 16.3 | 104 |
| λQ33Y[15] | 23.6 | 65.2 | 0.8 | 22.6 | 80 |
| λsA49G[15] | 33.2 | 56.9 | 1.06 | 20.7 | 80 |
| λD14A[15] | 44.0 | 61.0 | 1.48 | 21.3 | 80 |



**Figure 3.** Number of micro-routes $n$ vs a topological property of the native structure, the RCO, for each protein. The number of pathways correlates with the relative contact order. Fast-folders, which are typically helical proteins, have more micro-routes to native than slower folders. The correlation coefficient $R$ is 0.81



**Figure 4.** Flory interaction parameter (in kcal/mol) that best fits the rate data for each protein correlates with the chain length. Correlation coefficient is 0.91

number of routes $n$; the dimensionless chain monomer interaction energy $e$ related to Flory interaction paramter $\chi$ (see appendix A in the Supporting Information); and the barrier energies for folding $\epsilon_{DN}$ and unfolding $\epsilon_{ND}$. We note that using a more traditional strategy, based on an Eyring model,[57] would require instead a total of 6 parameters, and would give less information about the shape of the energy landscape. Moreover, an Eyring model would predict (incorrectly) that the folding rate goes to zero at infinite temperature,[58] in strong disagreement with the data.

Described below, the model predicts that the rates of folding and unfolding are proportional to $n$, the number of microscopic pathways. In addition, the model shows that the folding rate is reduced in proportion to the size of the denatured state space, which grows with temperature.

## III. Results: Fast-Folding Proteins Have More Micro-Routes than Slower Proteins

For the 13 proteins for which data is currently available, this model captures well the temperature dependences of the folding and unfolding rates (Figure 2). We have also fitted the overall relaxation rates ($k_r$); these are the quantities that are observed directly in experiments. Our model accounts for the crossover from Arrhenius-like positive activation behavior at low temperatures to anti-Arrhenius negative activation behavior at high temperatures, as follows. At low temperatures, folding is rate-limited by traditional barrier-crossing events along thruway routes to the native state. This is the positive activation region. In contrast, at high temperatures, increasing the temperature expands the denatured chain, leading to a larger search space ($m(L, T)$ increases), which slows the search the chain must make through its dead-ends to find thruways to the native state (see eq 8).

Table 1 gives the fit parameters for each of 13 ultrafast folders, and for 2 slower-folding proteins, L9, and apocytchrome. (We have excluded the peripheral subunit protein because of the significant scatter in the experimental kinetic data (see Figure 2).) What can we learn from the values of the
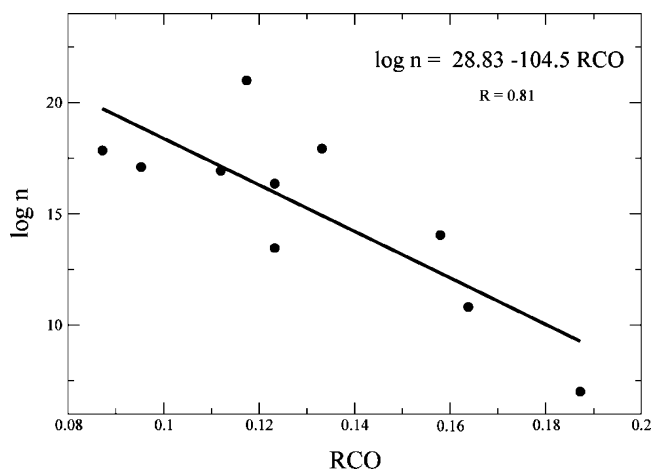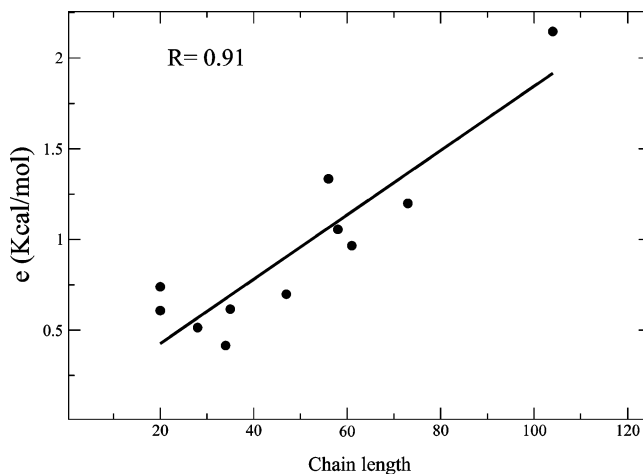
(57) Glasstone, S; Laidler, K. J.; Eyring, H. *Theory of Rate Processes*; McGraw-Hill: New York, 1940.
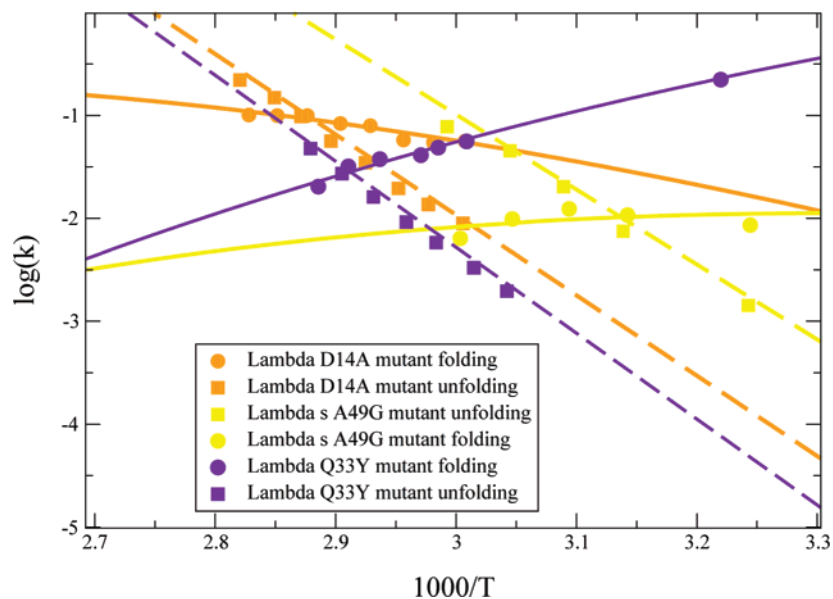(58) Ibarra-Molero, B.; Makhatadze, G. I.; Matthews, C. R. *Biochemistry* **2001**, *40*, 719−731.

fit parameters? First, Figure 3 shows that the number of routes, $n$, found in the model correlates with a structural property of the native state, the relative contact order (RCO). That is, the fastest-folding proteins (mainly helical) have the largest multiplicity of folding routes; α-β and β proteins have fewer folding routes. There are more folding routes for helical proteins because there are more places in the sequence where a helix can nucleate its folding. In β structures, nucleation is dominated by starting at a single site, namely the turn.

Second, Figure 4 shows that the best-fit value of the interaction parameter $e$ increases with the chain length. Larger values of $e$ correspond to a greater driving force to fold. The correlation in the figure indicates a greater folding cooperativity of larger proteins.

The model predicts that larger proteins have more compact denatured states, hence smaller values of $m$, hence their rate-limiting step is not conformational searching. Hence these slower larger proteins all have barrier-limited kinetics, that is, positive-activation Arrhenius kinetics, for all temperatures, consistent with experiments.

An interesting experimental system is the set of different mutants of λ-repressor fragment $λ_{6-85}$.[15] It is interesting

**Figure 5.** Model fits to folding and unfolding rate (in the units of $\mu$ s$^{-1}$) data for different mutants of lambda repressor protein.[15] (●) Folding rates; (■) unfolding rates. Solid lines are theoretical fits to the folding rates and broken lines are the same for the unfolding rates predicted from the model. We follow the same color scheme used in the original paper.[15]
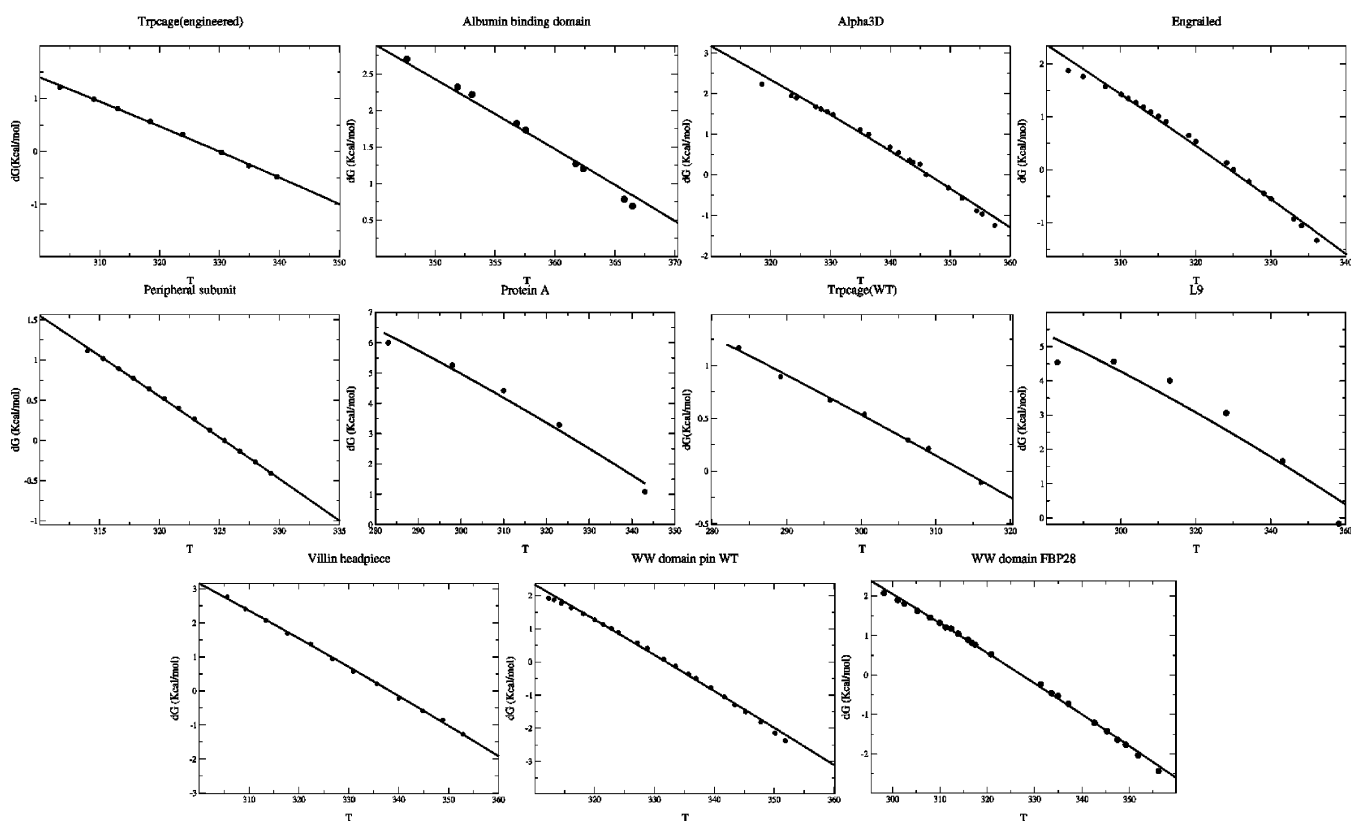


**Figure 6.** Model fits to temperature dependence of free energy of folding ($\Delta G_{DN}(T)$) for 11 proteins. (●) Experimental data; solid lines indicate the theoretical fits from the model.

because these otherwise identical proteins have different folding rates with different temperature dependences. We have fitted our model to the data on three mutants, Q33Y, A49G, and D14A (see Figure 5). In our model, Q33Y has the largest number of folding pathways and A49G has the smallest number (see Table 1), consistent with their corresponding folding speeds: Q33Y is fastest and A49G is slowest. The large value of *e* for D14A implies a compact denatured state, so the conformational search

is not limiting, and thus Arrhenius behavior is predicted at all temperatures. At the other extreme, Q33Y has the smallest value of *e*, meaning that the largest conformational search through the most expanded denatured state, which implies that rate limit is the diffusive search, hence non-Arrhenius kinetics.

Any model that captures the folding kinetics vs temperature should also capture the folding equilibrium vs temperature. Figure 6 confirms this for the 11 proteins for which equilibrium
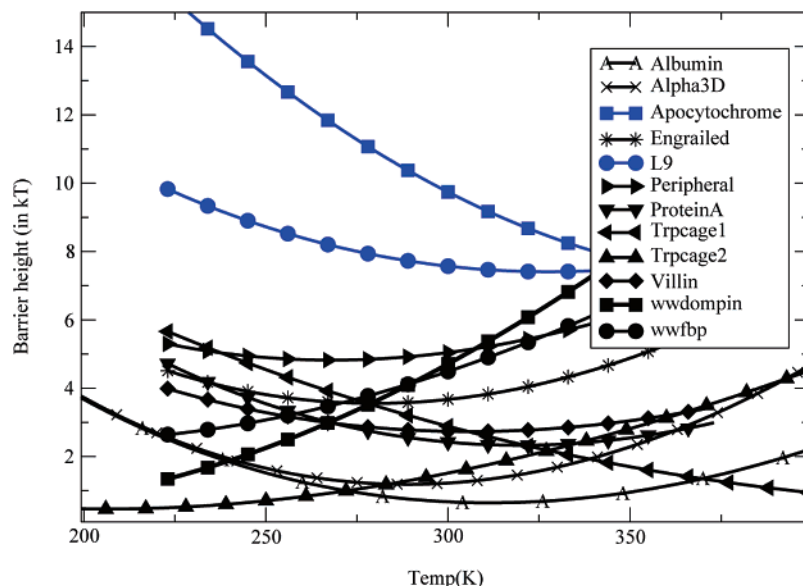
**Figure 7.** Free energy barrier heights ($\Delta G^{\ddagger}(T)$) for different proteins at different temperatures (from eq 13). The two slower folding proteins, Apocytochrome and L9, are shown in blue.

folding free energies are available. As an aside, for fitting equilibrium stabilities, an even simpler model is to use the difference in barrier heights $\Delta\epsilon = \epsilon_{ND} - \epsilon_{DN}$ and $e$ as two fitting parameters, in contrast to the more usual practice of fitting the folding free energy with three parameters: $\Delta C_p$, $\Delta S(T_m)$, and $T_m$, for example. The details are given in appendix D in the Supporting Information. The parameters obtained using this simpler 2-parameter model are consistent with the parameters from our kinetic model above.

## IV. Are There Energy Barriers to Ultrafast Folding?

Ultrafast folding has been called downhill or barrierless[5,8,17] because of the observation that the kinetics is sometimes non-exponential. However, in other cases, single-exponential kinetics is observed, for example in the engineered $\lambda_{6-85}$ protein, monitored using IR or fluorescence probes under strong native conditions.[18] Munoz and co-workers concluded that BBL is a downhill folder based on studies of melting temperatures, large changes in heat capacity, and continuous transitions observed by NMR during thermal unfolding.[42-46] However, studies by Ferguson et al.[49] on three bacterial homologues of that protein show single-exponential kinetics and a cooperative unfolding denaturation.

Here, we propose an independent way to test our model. The Thruway Search model predicts dynamics that is single-exponential at high and low temperatures, and with a possibility of a more complex kinetics in between, depending on the protein. Our model predicts a folding time distribution (Equation B6, Appendix B in the Supporting Information). In principle, this distribution could be observed in single-molecule folding experiments,[59,60] when such experiments become available. In particular, we can calculate the variance of folding times from the second moment ($\langle\tau^2\rangle$). We divide the variance by $2\times$ the square of the first moment ($2\langle\tau\rangle^2$), a property introduced and called $R_2$ by Onuchic et al.[61] (see details in the Appendix C in

the Supporting Information). As derived in equation C6 (appendix C in the Supporting Information) we find

$$R_2 - 1 = \frac{q k_1 m_a}{k_2 m^2} \tag{11}$$

This quantity $R_2$ is a useful measure of the non-exponentiality of the rate distribution. If a protein folds with ideal two-state single-exponential kinetics, then $R_2 = 1$. Our model predicts a single exponential ($R_2 \rightarrow 1$) at low temperatures, because the trap barriers are much smaller than the folding barriers at those temperatures ($k_1/k_2 \rightarrow 0$ in eq 11). Our model also predicts a single-exponential at high temperatures because $R_2 \rightarrow 1$ as $m$ increases ($m_a/m^2 \rightarrow 0$ in eq 11). For midrange temperatures, however, for some values of parameters, it is possible for $R_2$ to be significantly greater than unity, meaning a large variance in rates, and thus non-exponential behavior. Thus, our model allows the possibility of both single and non-exponential behavior over the midrange of temperatures for different proteins. This may account for apparent contradictions in some experimental data. Using predicted quantities such as these, single-molecule experiments could provide useful tests.

In our model, each micro-route has an energy barrier. What is most relevant for comparing to experiments, however, is the height of the macroscopic *free energy* barrier. We explicitly assume the Arrhenius law (eq 12) since this is the standard way that macroscopic barriers are traditionally defined. The folding rate $k_f$ is expressed in terms of the folding free energy barrier, $\Delta G^{\ddagger}$ as

$$k_f(T) = k_0 \exp(-\Delta G^{\ddagger}(T)/kT) \tag{12}$$

where we have explicitly indicated the dependence on temperature. Comparing this with eq 8 gives

$$\Delta G^{\ddagger}(T) = \epsilon_{DN} - kT \ln\left(\frac{n}{m(L, T)}\right) \tag{13}$$

(59) Rhoades, E.; Gussakovsky, E.; Haran, G. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*(6), 3197−3202.
(60) Rhoades, E.; Cohen, M.; Schuler, B.; Haran. G. *J. Am. Chem. Soc.* **2004**, *126*, 14686−14687.

(61) Leite, V. B. P.; Onuchic, J. N.; Stell, G.; Wang, J *Biophys. J.* **2004**, *87*(6), 3633−3641.
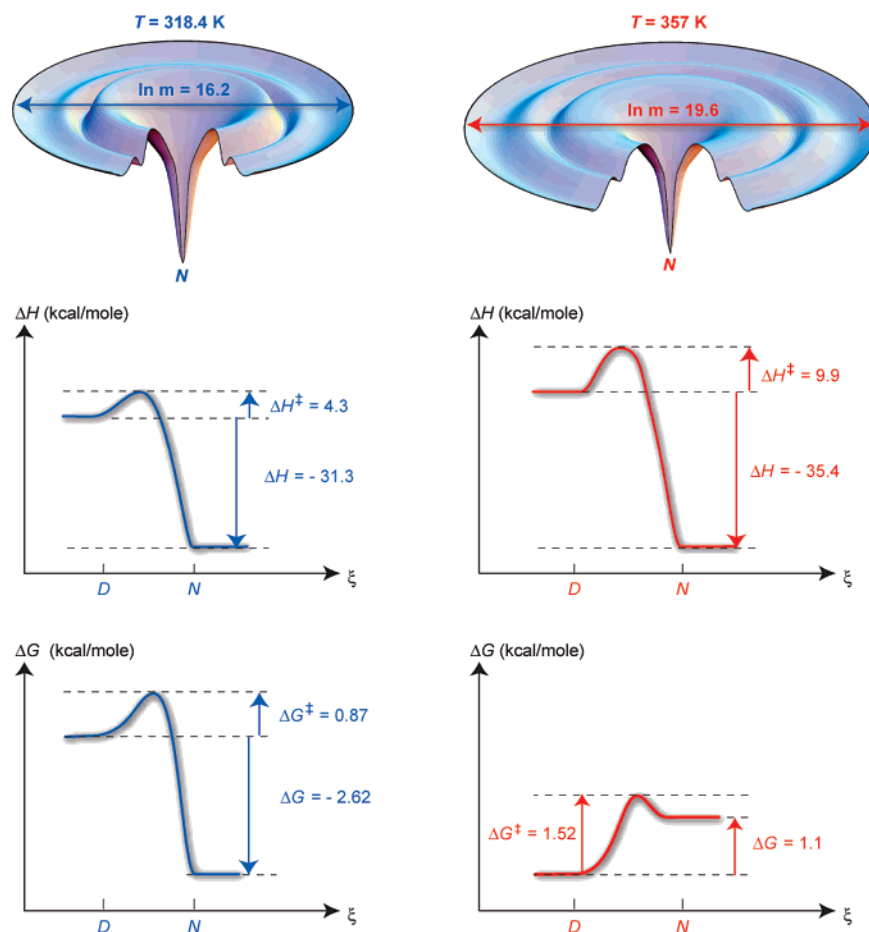
**Figure 8.** Interpreting the model, in two ways: with energy landscapes and reaction coordinate diagrams, at two different temperatures, 357 and 318.4 K. At low temperature, folding is rate-limited by energy barriers on the thruway micro-routes from $\mathscr{D}_0$ to $N$. At high temperatures, those energy barriers are readily overcome, and the rate-limit then becomes the speed the chain can search through its denatured trap states to find a thruway route to native. At the low temperature, the folding free energy is $\Delta G_{DN} = G_N - G_D = -2.62$ kcal/mol, enthalpic change $\Delta H_{DN} = -31.3$ (kcal/mol) and the change in entropy $\Delta S_{DN} = -44.25$ (in the Boltzmann unit, $k_B$). At the same temperature, we also calculate the barrier height, $\Delta G^{\ddagger} = 0.87$ (kcal/mol) whereas the enthalpic contribution to the barrier is $\Delta H^{\ddagger} = 4.3$ (kcal/mol) and the entropic contribution is $\Delta S^{\ddagger} = 5.3$ ($k_B$). Similarly, we calculate all the stability values and barrier heights at the highest temperature, 357 K. We report $\Delta G_{DN}(357) = 1.1$ (kcal/mol), $\Delta H_{DN}(357) = -35.4$ (kcal/mol), $\Delta S_{DN}(357) = -50.30$ ($k_B$), $\Delta G^{\ddagger}(357) = 1.52$ (kcal/mol), $\Delta S^{\ddagger}(357) = 11.5$ ($k_B$), and $\Delta H^{\ddagger}(357) = 9.9$ (kcal/mol). These numbers have been calculated from our model by fitting the data for α3D.

where we use the values of $n$ and $\epsilon_{DN}$ that are obtained from the curve-fits of experimental data and where $m(L, T)$ is obtained from equation A6 (see Supporting Information) based on the value of $e$ also obtained from the fit.

Our model indicates that for the ultrafast folders studied here, the free energy barriers can be very small (ranging from 0.5 kT for albumin binding domain, the fastest folder with the smallest barrier, to 10 kT for apocytochrome, a slow folder), but none of the barriers is truly zero (see Figure 7). These proteins all have positive free energy barriers at all temperatures (see Figure 7). The barriers themselves are temperature dependent. Munoz et al. have recently developed an NMR technique that can observe barriers as small as 3 kT;[46] this may prove quite useful for proteins of the type we have studied here. We note the caveat that our estimates of these barrier heights are based assuming the intrinsic rate, $k_0 = 100/L \, \mu \, \mathrm{s}^{-1}$.[4] This value of $k_0$ is an upper limit, meaning that our model could over-estimate the barrier heights.

Figure 8 is a summary of the model predictions, at two different temperatures. First, it shows a cartoon energy landscape, illustrating the different sizes of the search space at the two different temperatures. This indicates the role of the chain entropy in the folding kinetics. Second, the figure also shows

reaction coordinate diagrams, but with the enthalpy and the free energy shown separately here. Comparing these diagrams shows that enthalpy barriers can be substantial even when the free energy barriers are quite small. In addition, it shows that at high temperatures—above the denaturation transition—the enthalpy stabilizes folding, whereas the chain entropy destabilizes it.

Overall, at low temperatures, folding is dominated by traditional energy barriers along the thruway routes to the native state. At high temperatures, where those barriers are readily overcome, the rate-limit becomes the search of the chain through its trap states to find direct thruway routes to the native state. As the temperature increases, for ultrafast folders, the number of dead-ends increase (shown in Figure by increased size of the $\mathscr{D}_m$ basin), accounting for the negative activation behavior at high temperatures.

## V. Conclusions

We have developed a model of protein folding kinetics that combines the microscopic energy landscape funnel perspective with a macroscopic master-equation. Our goal was a model that could fit experimental data and where the parameters have some meaning in terms of the shape of the energy landscape of the microstates. We apply the model mainly to ultrafast folders,

where extensive thermal data is available. Unlike Eyring-based models, which would require six parameters and give no information about the energy landscape, the present model requires only four parameters to fit each protein and gives information about the shapes of the energy landscape. The model indicates that the fastest folders are proteins having a high multiplicity of parallel microscopic folding routes, and that the barriers for the fastest folders at high temperatures are due to the speed that the chain searches its conformations to find routes to the native state. It follows that these proteins have activation barriers at low temperatures and negative activation at high temperatures. The model also predicts folding time distributions;

these could give a definitive test of the model when single-molecule folding experiments become available.

**Supporting Information Available:** Details of our calculation of the number of dead-end states, effective folding rate, second moment, entropies, and enthalpies for folding equilibria are given. This material is available free of charge via the Internet at http://pubs.acs.org.

JA066785B