Title
Rate-limiting steps in protein folding

Permalink
https://escholarship.org/uc/item/61m6j8xm

Author
Schonbrun, Jack,

Publication Date
2001

Peer reviewed|Thesis/dissertation

Rate-Limiting Steps in Protein Folding

by

Jack Schonbrun

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of
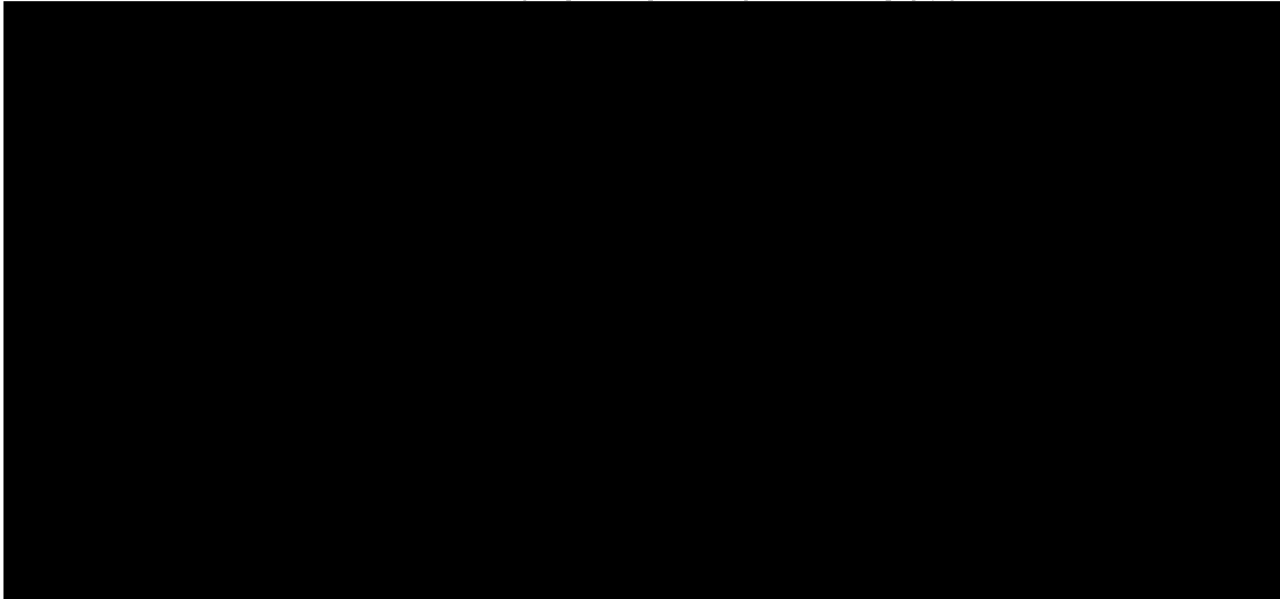
DOCTOR OF PHILOSOPHY

in

Biophysics

in the

GRADUATE DIVISION

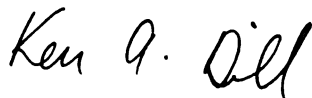of the

UNIVERSITY OF CALIFORNIA SAN FRANCISCO

Date   University Librarian

Degree Conferred: . . . .

Authorship

Chapter 1 includes Ken A. Dill as a co-author. Jack Schonbrun was the principal researcher for this study; the co-author directed and supervised the research.

Chapter 2 includes Hue Sun Chan and Ken A. Dill as co-authors. Jack Schonbrun was the principal researcher for this study; The co-authors directed and supervised the research.

Ken A. Dill

Ken Dill

Department of Pharmaceutical Chemistry

Abstract

Rate-Limiting Steps in Protein Folding

by

Jack Schonbrun

Doctor of Philosophy in Biophysics

University of California at San Francisco

Ken A Dill, Chair

Many proteins fold and unfold with single-exponential (2-state) kinetics, implying the existence of a transition state. What are the transition state conformations of a folding protein? We have solved the complete folding kinetics of a simplified protein model using a rigorous eigenvector method. We conclude that there is an alternative to barrier-based models for explaining 2-state folding. While Transition-State theory explains rates in terms of either energy or entropy barriers, the funnel model instead explains kinetics in terms of *entropic acceleration*. Single exponential kinetics can arise from the multiplicity of routes downhill to the native state on funnel- shaped landscapes. The rate of conformational diffusion at the top of the landscape is faster than intrinsic bond rotation rates because of the multiplicity of routes, and slows toward the bottom of the funnel, leading to a separation of time scales (fast at the top,

slowest at the bottom), and to single-exponential kinetics. Model predictions agree with experiments in giving a 'chevron' dependence of folding rate with the strength of native contacts, one of the main experimental fingerprints of 2-state folding.

Contents

List of Figures

Chapter 1

Transition States in Protein Folding

Jack Schonbrun and Ken A. Dill

Abstract. *Proteins are complex molecules, having thousands of degrees of freedom. Why is protein folding often simple, involving only single-exponential (called two-state) kinetics? Such kinetics is usually interpreted using transition-state theory, in terms of a reaction coordinate and an activation barrier, either energetic or entropic. Using a microscopic model, we find a different explanation: a rate* acceleration, *due to the funnel shape of the energy landscape, rather than a* deceleration *due to a bottleneck. Folding is fast relative to the microscopic transition rate because of multiple parallel reaction coordinates. The* apparent transition state—*the ensemble of rate-limiting microsteps—is the same ensemble as the denatured conformations, but with different statistical weights.*

The kinetics of protein folding is often remarkably simple. For many proteins, both folding and unfolding processes are single exponential functions of time *(1-5)*. And the ratio of the forward to reverse rate constants equals the equilibrium constant. Such situations can be described in terms of a two-state model,

$$D \leftrightarrow N,$$

and a corresponding Arrhenius diagram (Figure 1). Arrhenius diagrams originated to explain why chemical and physical processes are slower than a "speed limit," $kT/h = 0.16$ psec, where k is Boltzmann's constant, T is temperature, and h is Planck's constant. The main idea embodied in such diagrams is the concept of a bottleneck,

called the *transition state* or *activated state* *(6)*. The same idea is embodied in the Kramers' theory for reactions in solution *(7)*.

But protein folding is quite different than chemical reactions. The challenge is to explain why protein folding is so fast, not why it is slow. Hundreds to thousands of bonds must rotate for a typical denatured conformation to become native, and each rotation probably requires 1–100 psec, yet some proteins can fold in as little as 10 μsec *(2)*. There cannot be many mistakes or large barriers. Also, since each denatured conformation is so different than every other one, there cannot be a single microscopic trajectory that leads to the folded state.

Theoretical models have provided a microscopic understanding of several aspects of protein folding kinetics *(12-19)*. Two-state protein folding is now often described in terms of funnel-shaped energy landscapes *(20-22)* (see Figure 1b). Deeper on the landscape represents lower energies, and the width represents the number of conformations having a given energy. There are many unfolded conformations (high on the landscape), and very few low-energy native-like conformations (low on the landscape). This is mainly due to excluded volume: there are relatively few ways a polymer chain can configure into compact low-energy conformations *(23)*.

But it is not yet clear how two-state folding kinetics can result from a funnel-like energy surface, which has no apparent bottleneck. Put differently, what microstates (conformations) collectively constitute the macrostate called the transition state (TS), and how do they differ from those that constitute the denatured state
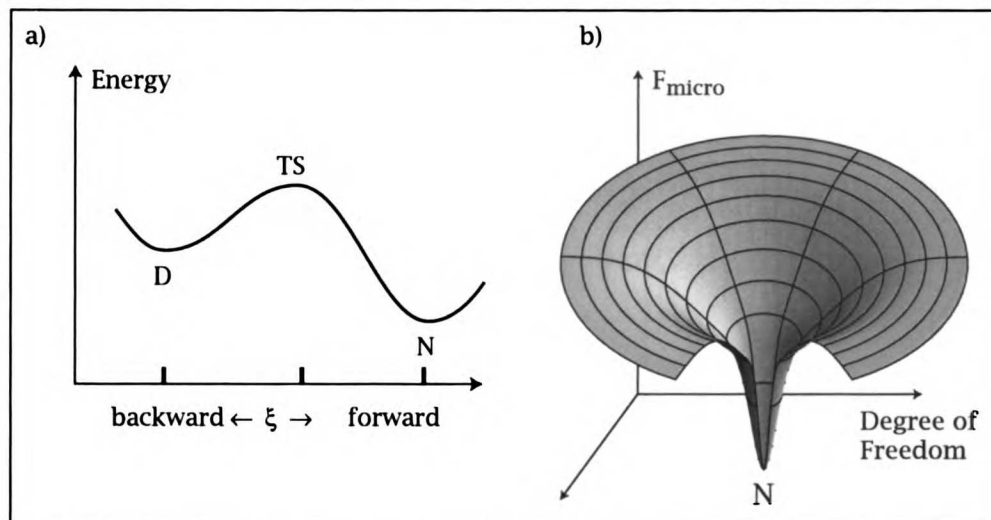
Figure 1.1: Transition-state explanation of single- exponential processes, such as protein folding, involves a rate- limiting step, shown as an obligatory thermodynamic barrier. **b,** Theory and simulations show that energy landscapes for protein folding are funnel-shaped, having no apparent microscopic energetic or entropic barriers.

(D) (8-11) ?

To address this question, we use a Gō model *(24,25)*. Gō models are widely used for studying two-state protein folding *(26-29)* because such models have two-state folding and unfolding kinetics, unique lowest-energy native states, exponentially large conformational searches with chain length, and because full landscape characterization is not yet computationally feasible for more detailed models. For the kinetics problem of interest here, we needed a rigorous and general way to identify the microscopic rate-liming conformations. We use a master equation method *(30)*.

We reach two main conclusions about how two-state protein folding differs from the traditional transition-state interpretation.
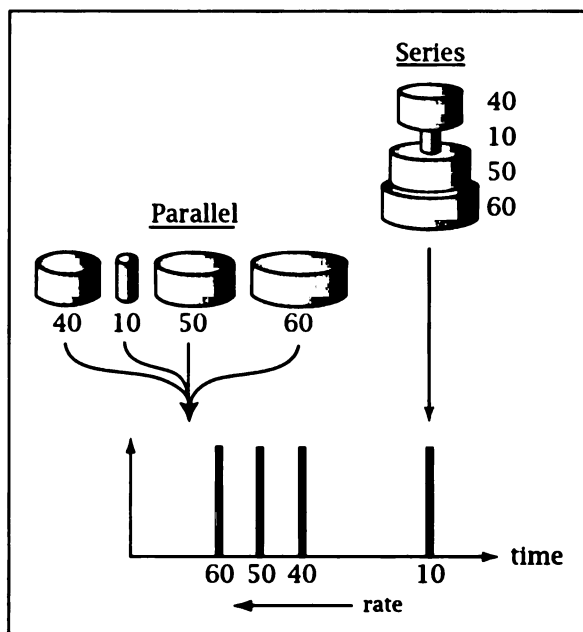
Figure 1.2: Two different models of microscopic kinetics. **Series:** When a macroscopic flow process results from microscopic fluxes in series, the macroscopic rate is limited by the microscopic bottleneck rate. The hypothetical numbers illustrate different microscopic step flow rates. **Parallel:** When microscopic flows are in parallel, the macroscopic rate is faster than the fastest microscopic rates.

(1) Parallel Routes Accelerate the Flow. Figure 2a shows the principle of a bottleneck: if microscopic flow processes are in *series*, one step will be rate-limiting. The rate of native structure formation would be limited by a particular conformational transition or the formation of a particular substructure of the protein. Figure 2b shows an alternative model. In a *parallel* flow process, the macroscopic flow rate is *faster* than the individual microscopic rates. Figure 3 shows that the the overall rate of folding in our model is faster than almost all of the microscopic conformational transition rates, hence most closely resembles the parallel flow model.

Why is folding so fast? The explanation is *entropic acceleration*. A previous

argument ('the Levinthal paradox' *(31)*) suggests that the folding search for the native structure should be slowest at the earliest stages because the chain has so many high-energy (open, denatured) conformations to explore. Our model shows the opposite: the chain wastes the least amount of its folding time in the high-energy states. The large numbers conspire to speed up folding, not to slow it down. If the intrinsic rate of transition from conformation i to j_0 is k_{i,j_0}, folding can happen faster than this because conformation i could also instead transition to any other conformation $j = 1, 2, 3, \ldots, j_0, \ldots, M$ that is downhill in energy. This multiplicity of routes downhill accounts for why the energy level-to-level rates are faster than the conformation-to-conformation rates. While transition-state theories invoke an intrinsic *maximum rate kT/h* and a *slowing step* (an energy barrier), the funnel model explains the high speed of folding in terms of an intrinsic *minimum rate* and *an acceleration* due to the multiplicity of microscopic routes to the native structure.

Single-exponential dynamics is usually taken to imply a relatively unique microscopic trajectory. How can multiple parallel routes explain the single-exponential dynamics observed in this model and in experiments? Single-exponential kinetics also occurs if early steps are fast enough to cause the last steps to be rate-limiting. On funnel landscapes, folding is fast throughout the high energy levels on the landscape, but much slower at the end, where the multiplicity of routes is smaller.

(2) The Transition State Conformations are Denatured Conformations

What are the rate-limiting conformations in the funnel model? There are 4 macrostates
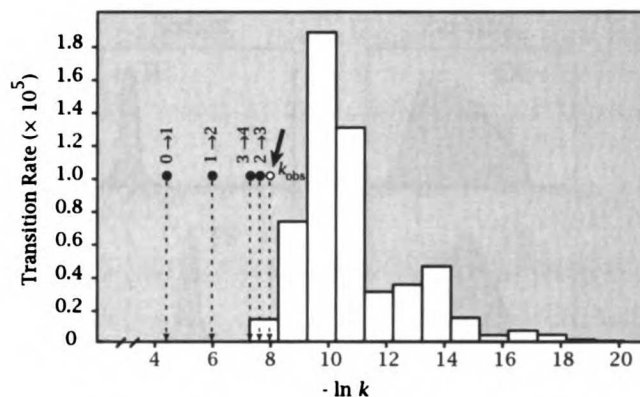
Figure 1.3: The rate of folding is faster than almost all the microscopic transitions between conformations. The large arrow shows the overall folding rate. The histogram to the right shows the microscopic transition rates. The small arrows on the left show the level-to-level transition rates between energy levels: 0 to 1 indicates transitions at the top of the energy funnel, 1 to 2, is the next level down, etc.

relevant in the model: N, the single native conformation, D_u, the Boltzmann-weighted denatured ensemble under the unfolding conditions from which folding is initiated, D_f, the Boltzmann-weighted non-native ensemble under particular folding conditions, and TS_f (the apparent transition state), the eigenvector of conformations that fold to the native state with the single-exponential relaxation time given by the slowest eigenvalue, under folding conditions.

In transition-state theories, the macrostates (Reactant, Intermediate, Transition State, Product) are all identifiable as distinct ensembles of microstates (specific atomic structures) that occur along a single reaction coordinate. Along any such one-dimensional coordinate, a state A is either 'before', or 'after' some other state B. Applied to proteins, it implies the transition state is *between*—and *distinct from*—the denatured state (D) and the native state (N): $D \rightarrow TS \rightarrow N$.
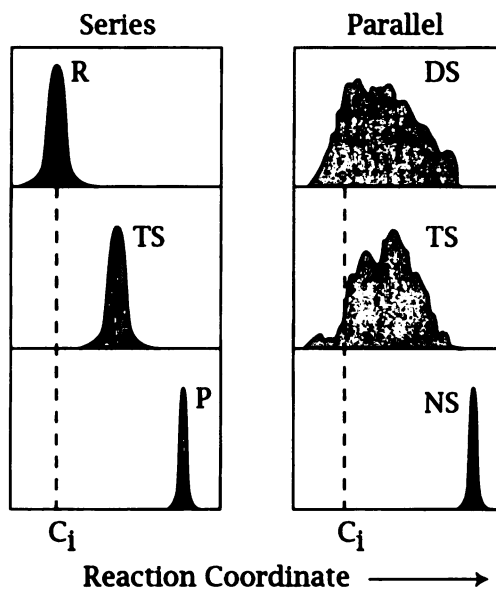
Figure 1.4: Series: the transition-state model involves macrostates (reactant, transition state, product) that are small localized ensembles of microstates. Each microscopic structure can be identified as being a member of only one macrostate. Parallel: the present model shows that the macrostates for two-state protein folding (D, the denatured state, or TS_f, the transition state under folding conditions) are broad ensembles that encompass all the non-native chain conformations. A given chain conformation is not a unique member of a macrostate.

But Figures 4 and 5 show that our folding model is at variance with these aspects of classical rate theory, in two respects. The right side of Figure 4 illustrates schematically how the TS_f ensemble can be so broad that it includes all the same conformations that are in D_u and D_f. Figure 5 shows the distributions from the model simulations. No individual chain conformation is exclusively in one state D_u, D_f, or TS_f without being in the others at the same time. These ensembles differ only in the statistical weights of the individual conformations. The folding process is the evolution of a single ensemble, not a time sequence of distinct ensembles.

Second, the ensembles D_u, D_f, and TS_f change with external conditions. If we
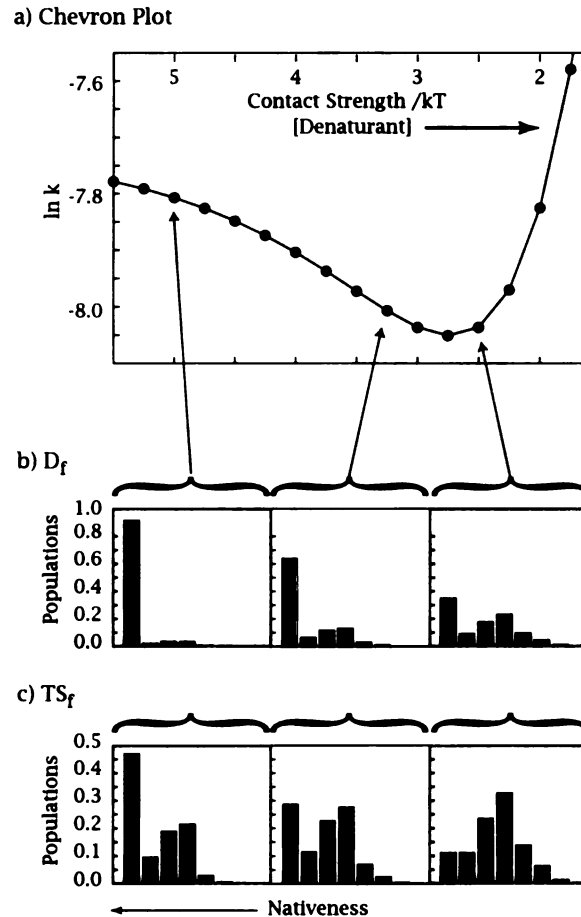
a) Chevron Plot

Figure 1.5: "Chevron" plot showing predicted folding rate *versus* strength of native contacts. The histograms show the distributions of conformations (most native-like to the right) under different folding conditions for the denatured ensemble D_f (bottom row) and the apparent transition state, TS_f (top row). As conditions become stabilizing of the native state, both ensembles D_f and TS_f become more native-like. Because of this, the left (folding branch) of the chevron curve shows "rollover", a deviation from linearity. Interestingly, under strong native conditions, the D_f ensemble is more native-like than the TS_f ensemble.

were to attempt to rank order these macrostates in terms of their structural similarity to the native state, figure 5 shows that under strong native conditions the order would be $TS_f \rightarrow D_f \rightarrow N$. This nonsensical result is a consequence of using a single reaction coordinate and a series model to represent a process that is intrinsically parallel.

The principal test of a two-state folding model is whether it predicts the 'chevron' (V- shaped) dependence of the folding rate on the driving force (denaturant or temperature) that is observed in experiments *(32–35)*. Figure 5 shows that this model does so. Moreover, some experiments show small downward curvature on the left (folding) branch of the chevron plot *(36, 37)*. This is also predicted by the model. What is the basis for this curvature? Classical theories expect no curvature, based on assuming that the TS doesn't change with external conditions. In that light, the curvature has been explained as resulting from intermediate states *(37)* or broad transition states *(38)*. But Figure 5 gives a different explanation, namely that the apparent transition state ensemble changes with denaturant or temperature. TS_f reaches a limiting native-like distribution under strong native conditions.

Summary Many proteins fold and unfold rapidly with single-exponential (2-state) kinetics. Previous explanations have been based on transition-state assumptions that: (1) that macrostates (D, TS, N) are distinct and unique collections of identifiable microstates, (2) there is a single reaction coordinate, so the terms 'forward' and 'backward' have meaning, and (3) the rate of folding does not exceed the rate of some microscopic (energetic or entropic) bottleneck step.

The present work offers a different interpretation: (1) The macrostates—the denatured and apparent transition states—encompass all the non-native conformations, but each macrostate is a different distribution of microscopic populations, (2) there are multiple microscopic routes in a high-dimensional space, so a single microscopic

reaction coordinate is not sufficient, (3) Folding is an acceleration relative to an intrisic slowest rate, not a slowing down relative to an intrinsic fastest rate, and (4) the rate of folding can be faster than individual microscopic transition rates. In this model, the apparent rate-limit that gives single- exponential behavior arises because the last microscopic transitions are slow relative to earlier steps. Early steps are accelerated by the multiplicity of folding routes at the top of the funnel. In this view, the message in two-state folding kinetics is the accelerator, not the brakes.

Bibliography

[1] S. E. Jackson, *Fold. Des.* **3**, R81 (1998).

[2] R. E. Burton, J. K. Myers, T. G. Oas, *Biochemistry* **37**, 5337 (1998).

[3] P. S Kim, R. L. Baldwin, *Ann. Rev. Biochem.* **59**, 631 (1990).

[4] A. R. Fersht, *Structure and Mechanism in Protein Science.* (Freeman, New York 1999).

[5] W. Englander, *Ann. Rev. Biophys. Biomol. Struct.* **29**, 213 (2000).

[6] H. Eyring, *Chem. Rev.* **17**, 65 (1935).

[7] H. A. Kramers, *Physisa* **7**, 285 (1940).

[8] A. R. Fersht, *Current Biology* **5**, 79 (1994).

[9] V.P. Grantcharova, D. S. Riddle, D. A.. Baker, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 7084 (2000).

[10] T. R. Sosnick, L. Mayne, R. Hiller, S. W. Englander, *Nat. Struct. Biol.* **1**, 149 (1994).

[11] N. Taddei, *et al.*, *J. Mol. Biol.* **300**, 633 (2000).

[12] J. D. Bryngelson, J. N. Onuchic, N. D. Socci,P. G. Wolynes, *Proteins* **21**, 167 (1995).

[13] V. S. Pande, A. Yu. Grosberg, T. Tanaka. *Fold. Des.* **2**, 109 (1997).

[14] H. S. Chan , K. A. Dill. *J. Chem. Phys.* **100**, 9238 (1994).

[15] N. D. Socci, J. N. Onuchic, Peter G. Wolynes. *Proteins* **32**, 136 (1998).

[16] D. K. Klimov, D. Thirumalai, *J. Mol. Biol.* **282**, 471 (1998).

[17] E. I. Shakhnovich, V. I. Abkevich, O. B. Ptitsyn. *Nature* **379**, 96 (1996).

[18] Y. Zhou, M. Karplus, *Nature* **401**, 400 (1999).

[19] K. A. Dill. *Prot. Sci.* **8**, 1166 (1999).

[20] P. E. Leopold, M. Montal, J. N. Onuchic, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 8721 (1992).

[21] P. G. Wolynes, J. N. Onuchic, D. Thirumalai, *Science* **267**, 1619 (1995).

[22] K. A. Dill, H. S. Chan, *Nat. Struct. Biol.*, **4**, 10 (1997).

[23] H. S. Chan, K. A. Dill, *Macromolecules*, **22**, 4559 (1989).

[24] H. Taketomi, Y. Ueda, N. Go, *Int. J. Pept. Prot. Res.* **7**, 445 (1975).

[25] We use a two dimensional lattice formulation of the Gō model. The conformation-to-conformation transition rate is given by $k_{ij} = \exp(-B\theta - \eta r)$, where θ is the number of dihedral angles that differ between structures i and j, and r is the sum of the squares of the displacements of each bead of the chain. The parameters B and η correspond to the energies for dihedral angle rotation and diffusion through the solvent, respectively. Values of $B = 0.5$ and $\eta = 0.1$ give transition probabilities comparable to those in traditional Monte Carlo dynamics move sets. The transition rates are multiplied by a Metropolis factor: the lesser of $\exp(-H_{ij})$ or 1, where H_{ij} is the difference in energy between the first and second conformations. In a Gō model this energy is solely determined by the number of native contacts.

[26] S. Takada, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 11698 (1999).

[27] O. V. Galzitskaya, A. V. Finkelstein, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 11299 (1999).

[28] E. Alm, D. Baker, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 11305 (1999).

[29] V. Muñoz, E. R. Henry, J. Hofrichter, W. A. Eaton, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 11311 (1999).

[30] The evolution of the ensemble with time is governed by a master equation *(39,*

40) : $dp(t)/dt = \mathbf{K}p(t)$, where p is a vector that contains the population of each conformation at each instant, and $\mathbf{K}$ is a matrix that contains the transition probability (per unit time) between each conformation. The elements of this matrix are the weighted k_{ij}'s described above. The dynamical evolution of this entire ensemble can be solved exactly in terms of the eigenvectors and eigenvalues of the matrix $\mathbf{K}$, using readily available numerical algorithms *(41)*. The smallest non-zero eigenvalue corresponds to the rate constant for the onset of the native state. Its corresponding eigenvector represents all the relative contributions of the microscopic transitions that contribute to the slowest, rate- limiting step. Hence this defines the TS macrostate..

[31] C. Levinthal, *J. Chim. Phys.* **85**, 44 (1968).

[32] S. E. Jackson , A. R. Fersht, *Biochemistry* **30**, 10428 (1991).

[33] D. E. Otzen, O. Kristensen, M. Proctor, M. Oliveberg, *Biochemistry* **38**, 6499 (1999).

[34] R. E. Burton, G. S. Huang, M. A. Daugherty, T. L. Calderone, T. G. Oas, *Nat. Struct. Biol.* **4**, 305 (1997).

[35] C. R. Matthews, *Annu. Rev. Biochem.* **62**, 653 (1993).

[36] S. Khorasanizadeh, I. D. Petrs, H. Roder, *Nat. Struct. Biol.* **3**, 193 (1996).

[37] A. Matouschek, J. T. Kellis, L. Serrano, M. Bycroft, A. R. Fersht, *Nature* **346**, 440 (1990).

[38] M. Oliveberg, *Accts. of Chem. Res.* **31**, 765 (1997).

[39] H. S. Chan, in *Monte Carlo Approach to Biopolymers and Protein Folding*, P. Grassberger, G. T. Barkema, W. Nadler, Eds. (World Scientific, 1998) pp. 29–44.

[40] M. Cieplak, J. Karbowski, M. Henkel, J. R. Banavar. *Phys. Rev. Let.* **80**, 3654 (1998).

[41] W. H. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery. *Numerical Recipes in C* (Cambridge Univ. Press, New York 1992), pp 469–478.

[42] J. S. was a Howard Hughes Predoctoral Fellow.

Chapter 2

Entropic Acceleration: An Alternative to Transition-State Explanations of Protein Folding Kinetics

Jack Schonbrun, Hue Sun Chan and Ken A. Dill

Abstract. *Many proteins fold and unfold with single-exponential (2-state) kinetics, implying the existence of a transition state. What are the transition state conformations of a folding protein? We have solved the complete folding kinetics of a simplified protein model using a rigorous eigenvector method. We conclude that there is an alternative to barrier-based models for explaining 2-state folding. While Transition-State theory explains rates in terms of either energy or entropy barriers, the funnel model instead explains kinetics in terms of entropic acceleration. Single exponential kinetics can arise from the multiplicity of routes downhill to the native state on funnel- shaped landscapes. The rate of conformational diffusion at the top of the landscape is faster than intrinsic bond rotation rates because of the multiplicity of routes, and slows toward the bottom of the funnel, leading to a separation of time scales (fast at the top, slowest at the bottom), and to single-exponential kinetics. Model predictions agree with experiments in giving a 'chevron' dependence of folding rate with the strength of native contacts, one of the main experimental fingerprints of 2-state folding.*

2.1   Introduction: Of funnels and barriers

How does single-exponential protein folding and unfolding kinetics arise from funnel-shaped energy landscapes? The energy landscape of a protein is the mathematical function that gives the relative free energies of all its conformations. It gives the microscopic framework for understanding the thermodynamics of folding and of

all of a protein's conformational transitions. The landscape also underlies folding kinetics, since folding is a process of navigating from the non-native states high on the landscape downhill toward the native state.

Despite the complexity of a typical protein, experiments often show remarkably simple single-exponential folding or unfolding curves [1, 2, 3, 4, 5]. That is, when an ensemble of such protein molecules is jumped from unfolding to folding conditions, the native population follows a single exponential function of time, with a well-defined rate constant. The rate constant depends on the amino acid sequence, and on the folding conditions, but typically not on the initial state of the protein.

There are two important facts that must be addressed in any viable theory of how proteins fold. First, folding can often be described by simple mass-action schemes, involving only 2 or 3 states, such as $D \rightleftharpoons N$, or $D \rightleftharpoons I \rightleftharpoons N$ [1]. Second, proteins become native much faster than would occur by exhaustive searching through conformational space [6]. These two points have been addressed by two different pictures of folding: the barrier picture and the funnel picture.

To describe these two pictures, we first define some terms. A *microstate* refers to a single conformation of a protein molecule: every bond length and angle has a particular value. The list of all such microstates gives the *conformational space*. A *macrostate* refers to some particular collection (ensemble) of microstates, a subset of the conformational space. Examples of macrostates include the denatured state D; intermediate states I; the transition state TS. The native state N is a macrostate

that is so sharply defined as to be (approximately) identical to a single microstate.

The barrier picture developed around the idea that folding involves rate-limiting steps. When a physical process or chemical reaction has single-exponential kinetics in both forward and backward directions, it can be described in terms of two states in a mass-action model, $A \rightleftharpoons B$. Transition-State theory developed by Arrhenius, Eyring, and others [7] explains single-exponential kinetics in terms of a thermodynamic barrier between states A and B. Transition-state theory uses the concept of a *reaction coordinate*, a 1-dimensional degree of freedom where 'backwards' means toward the reactant and 'forward' means toward the product. The *reaction coordinate* is an example of a *macropathway*, an ensemble-averaged descriptor of progress. The term *micropathway* refers to the trajectory of a single chain as it folds. Whereas a reaction coordinate is 1-dimensional, a micropathway occurs in a high-dimensional space, the conformation space. Whereas progress happens as a monotonic function of time along a reaction coordinate, a micropathway can involve backtracking and random meandering; it is not monotonic. Whereas progress is simple to define along a 1-dimensional coordinate, there is often no unambiguous meaning of the word 'progress' in a high-dimensional space, since a conformational change can be forward along some coordinates and backward along others at the same time.

The power of transition-state theory to give microscopic insights into chemical reactions results because the reaction coordinate macropath coincides with essentially one (ensemble-averaged) micropath of the reacting molecules. This simplicity arises

because chemical bond energies are much greater than kT, so thermal variations from the dominant micropathway are relatively unimportant for chemical reactions, the traditional focus of Transition-State theory.

Two premises are needed to explain single-exponential kinetics in mass-action models. First, the macrostates (D, I, TS, for protein folding) are assumed to be time-independent (for example, the denatured state is the same distribution of microstates at the beginning of a folding experiment as at the end). Second, the interconversion rate between the two states is described by a rate 'constant' that is independent of time. Transition-State theory supposes that there is an equilibrium between the reactants and the transition state, and that the population of the rare transition state conformations are given by the Boltzmann distribution law. Much of the current discourse on protein folding revolves around the theoretical and experimental search to identify the microscopic conformations that comprise such a transition state [8, 9]. The idea is that identifying the slow steps will define the mechanism of protein folding.

But protein folding is rather different than simple chemical kinetics, and it is not so clear that the transition state concept applies. First, transition state theory describes an upper 'speed limit', kT/h, where k is Boltzmann's constant, T is temperature and h is Planck's constant ($kT/h = 0.2$ psec for $T = 300$ K). The point of transition-state theory is to explain why chemical reactions are slower than this speed limit. The explanation is that there is a free energy barrier.

But the key problem in protein folding is to explain why it is fast, not why it

is slow. A typical protein has hundreds of thousands of degrees of freedom, yet some proteins can fold within 10 μsec [10, 11]. Second, chemical reactions have well- defined reactants (R) and products (P) that are interconnected along a single microscopic reaction coordinate (describing specific structural changes). For proteins, the product (native state, N) is well-defined, but the 'reactant' is the ensemble of all non-native states, so there cannot be a single microscopic reaction coordinate from R to P. A key point is that the macroscopic reaction coordinate for folding does not coincide with any single microscopic trajectory. Protein folding from the denatured state D to the native state N is a *many* $\rightarrow$ *one* mapping, not a *one* $\rightarrow$ *one* mapping. Third, theoretical studies show that the shape of the protein folding energy landscape is funnel-like [12, 13, 14, 15, 16], for which there is no obvious energetic or entropic barrier.

If 2-state protein folding does not involve a transition-state barrier or a single microscopic reaction coordinate, how can we explain the single-exponential kinetics? The funnel picture captures the fact that there are many high-energy (open, unfolded) conformations of a protein and very few low-energy (compact, native-like) conformations . The funnel picture resolves the Levinthal "paradox", that a protein would not have enough time to find its native structure by a random search through conformational space. On a funnel, there is an energetic bias driving the protein downhill toward the more native-like structures. But a smooth funnel has no obvious rate-limiting bottleneck step - energetic or entropic - that could explain

single-exponential kinetics. A key question is how funnel-shaped energy landscapes could give rise to single- exponential kinetics, and what protein conformations constitute the rate-limiting step? This is the problem we address here.

We solve here a 2-dimensional Gō lattice model [17], in full detail. Lattice models have been a common method for modelling protein folding [18, 19, 20, 21] (See Dill, et al. [22] for a review.) Such models are widely used because they are simple enough that the full conformational space can be explored, yet they have the key features of 2-state protein folding [23, 24, 25, 26]: (1) a single native state in a conformational space that grows exponentially with chain length, (2) they include chain connectivity, attractions, and excluded volume, (3) they have 2-state folding and unfolding behavior. Conformations are shown in Figure 2.11.

To model folding, we start with a maximally denatured ensemble, which is the equilibrium distribution when the contact energy is equal to zero. We initiate folding by setting the contact energy to a value that favors the native conformation ($\varepsilon = -3.5kT$). We then follow the kinetics. Three different sequences all show the same behavior, so we focus only on a representative one here. As in previous studies [27], we find a wide range of time scales in the folding process. However, the emergence of the native state is dominated by a single rate constant, as is often seen in experiments (Figure 2.14.) There are many aspects of the folding process that take place on time scales faster than that of the onset of the full native conformation. Indeed, there are significant rearrangements of the distribution of non-native

conformations on a very fast time scale, sometimes called a 'burst phase'. But folding from that re-adjusted state appears then as a single exponential process because these rearrangements happen significantly more quickly than overall time scale of folding.

2.2   The Eigenvalue Perspective: The Distribution of Time Scales.

The present work is based on a general and rigorous way to identify the microscopic conformations that define the rate-limiting step in folding. Our method is different than traditional transition- state theory. Transition-State theory assumes: (1) macrostates and rate constants are time- independent, (2) the rate of product formation is proportional to a quasi-equilibrium population of transition-state molecules that are rare, hence described by a free energy barrier, and (3) there is a dominant micropathway that corresponds to the macroscopic reaction coordinate. Using further assumptions about how to define the two states (D and N) [9], Pande et al. give a way to identify transition state conformations. But such methods are not general enough to apply to folding on funnel landscapes. We sought a true kinetic method that would be rigorous, general, and not based on any of these assumptions, including even the thermodynamic premise.

Our approach is based on eigenvector solutions to the Master Equation [28]. The exact solution of the Master Equation of folding gives the full time course of folding.

In addition, we use the eigenvectors and eigenvalues of the **time-evolution** matrix to gain further insight into the origin of single exponential kinetics, and the values of rate constants for folding. These methods have the advantage that we can treat the whole ensemble of conformations. This is important here because we find that the essential features of the folding kinetics are not a property of the individual trajectories, or even ensemble averages over them. The present kinetics is characterized by the *number of trajectories*, and how that number diminishes down the energy landscape, rather than by the sizes of the barriers (energetic or entropic) along any one trajectory.

More specifically, the full conformational space is 740 conformations. The rate matrix that describes all the conformation-to-conformation transitions is 740 × 740. Therefore, 740 equations (in matrix form) describe the full time evolution of every conformation. We solve this first-order time- dependent matrix differential equation for its eigenvalues and eigenvectors. There is a dominant eigenvalue that rigorously accounts for all the conformational dynamics responsible for the slowest (single-exponential) step. The corresponding eigenvector gives the populations of all the microstates that are responsible for the slowest step. This then defines the 'apparent' transition state, i.e. the rate-limiting steps. We use the word 'apparent' here because we find that it is not a traditional transition state, and is not due to a barrier. The most striking result is that the apparent transition state is not distinct from the denatured conformations, as is usually assumed. Rather, the chain conformations that constitute the apparent transition state ensemble are the same microstates that

comprise the denatured ensemble, but they have different statistical weights.

2.3   Results

2.3.1   Distribution of Eigenvalues and Leaving Rates

For a given monomer sequence, we obtain a distribution of eigenvalues. The time course of the populations of conformation in the ensemble can be expressed exactly as a sum of exponentials, with rate constants that are inverses of the eigenvalues. Mathematically, there must be as many eigenvalues as conformations. Although degeneracy is possible, we do not find any in our model. Figure 2.1 shows the distribution of non- zero eigenvalues as a spectrum of rates for each of the sequences we studied, under strong folding conditions. (One eigenvalue must be zero, because at equilibrium there is no change, and hence one time scale of the ensemble must be infinite.) The eigenvalues are distributed over several orders of magnitude, meaning that folding involves a broad range of time scales. However, despite the existence of 740 relaxation times for this system, the rate of emergence of the native conformation is dominated by a single slowest non-zero eigenvalue, corresponding to single-exponential folding kinetics.

What is the meaning of these eigenvalues? In general, it can be difficult to rationalize eigenvalues, because each one is a linear combination of all N^2 elements in the matrix. However, here we can attribute a physical meaning to them. The eigenvalues
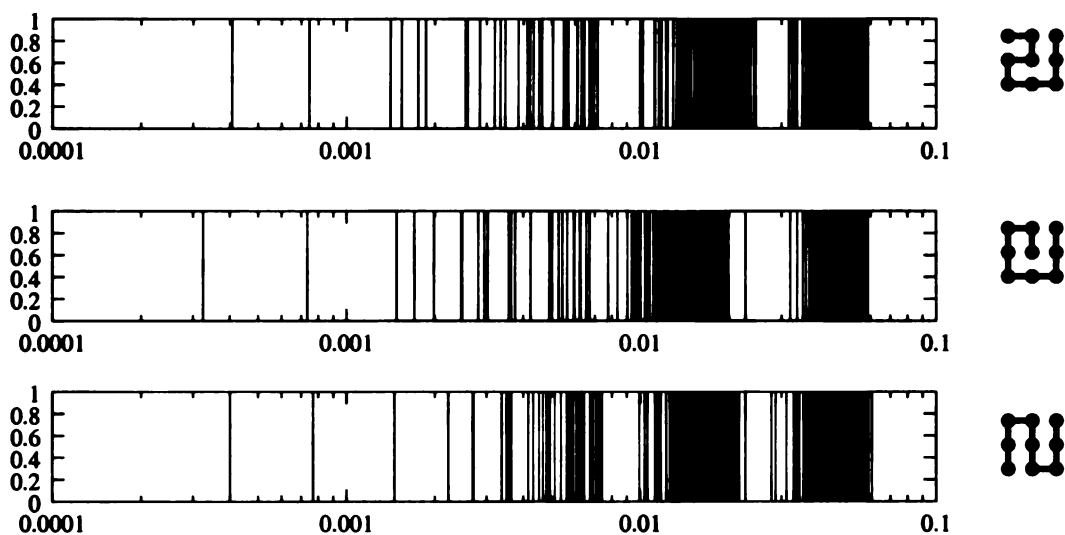
Figure 2.1: The spectra of non-zero eigenvalues for the three possible native structures of length nine. In each case there is the slowest eigenvalue is significantly slower than all others. The clustering corresponds to different energy levels of the landscape, as described in the text.

(which have units of inverse time) correlate directly with the *leaving rates*, the rates the protein molecules escape from the various conformations (figure 2.2.) Each leaving rate is the inverse of the mean lifetime of a conformation. For any conformation, this escape rate is the sum of all rates out to other conformations. These leaving rates are the diagonals of the matrix K in the master equation, as shown in equation 2.3. These diagonal elements are generally larger in magnitude than the off- diagonal elements, for the simple reason that the magnitude of each diagonal element is the sum of all the off- diagonal elements in its column. Physically, this means that that the leaving rate from a conformation is fast relative to individual conformation-to-conformation transition rates, since the leaving rate is a sum of all possible routes out. Since the diagonal elements dominate the matrix, they approximately equal the

eigenvalues, which, of course, are the diagonal elements of the diagonalized matrix.

We believe it may be a general principle that the master equation for a funnel landscape may often be a nearly diagonal matrix because of the high multiplicity of escape routes of most conformations at the top of the landscape.

Leaving rates are greatest at the top of the funnel: the most open conformations have the largest numbers of escape routes (mostly downhill). Deeper down the funnel, the chain is highly constrained, so only a few conformational transitions are viable escape routes, and those are mostly uphill. To see this, we express the leaving rates in terms of an effective free energy barrier. This barrier results from the ensemble of transitions out of a conformation. The weighting of each member of this ensemble is given by $e^{-\Delta H_{ij}^{\ddagger}/kT}$, where ΔH is given by equation 2.2. The free energy of the total barrier to leave a given conformation is proportional to $-kT \ln Z_i$, where Z_i is the partition function of the escape ensemble for conformation i:

$$Z_i = \sum_j e^{-\Delta H_{ij}^{\ddagger}/kT}.$$

Figure 2.3 shows that escape becomes increasingly difficult toward the bottom of the landscape (the free energy increases). The figure also shows the entropic and energetic components of this barrier. It shows that the enthalpic component of the barrier actually decreases with depth.

In contrast, the escape entropy becomes less favorable deeper in the landscape. This entropy is a measure of the number of exit routes out of each conformation. For conformations with few contacts, there are many ways out, resulting in short
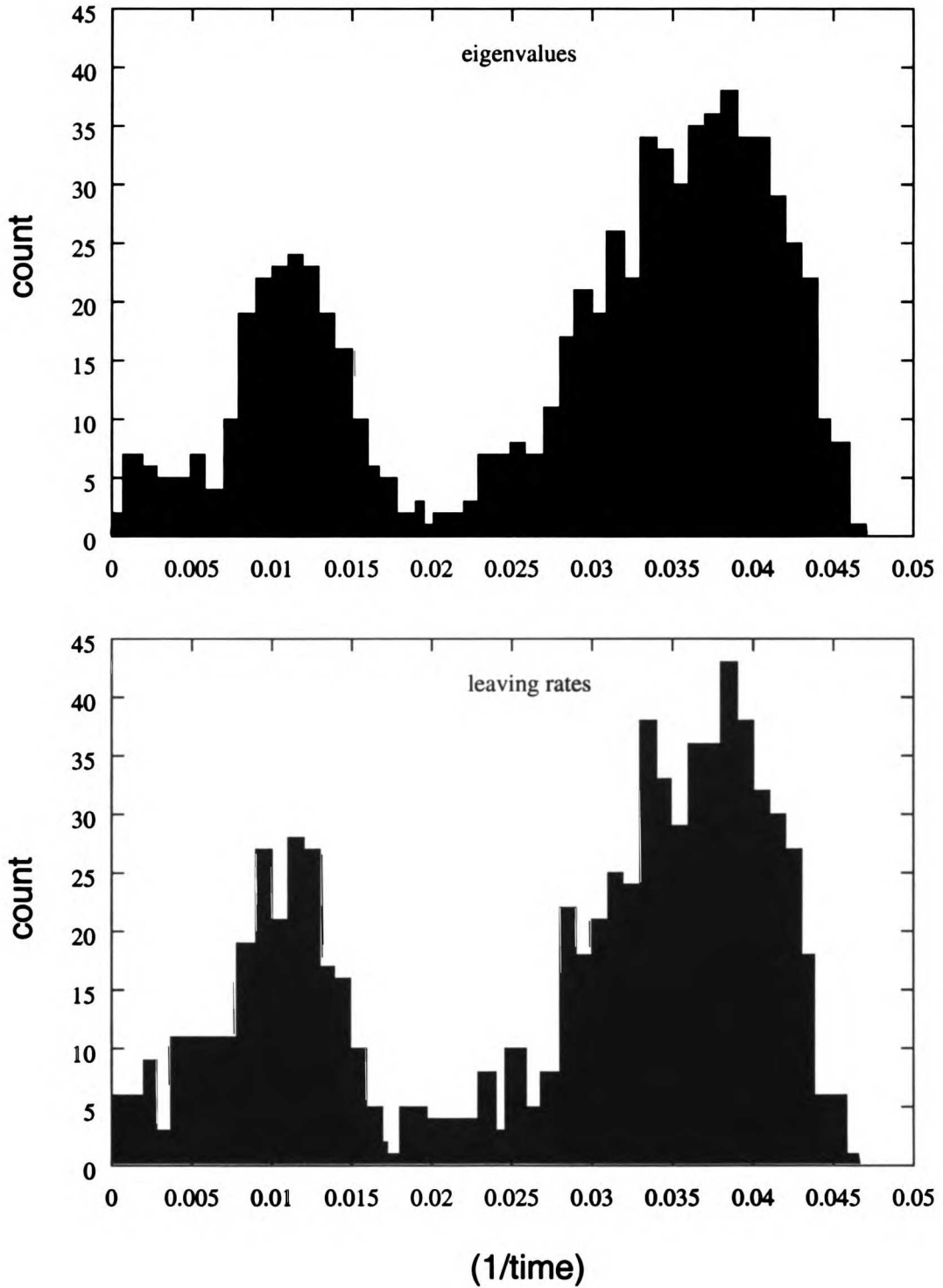
Figure 2.2: The eigenvalue distribution is well modeled by the distribution of leaving rates. These histograms are for the sequence ABCAXCXCB, results for the other sequences are similar.
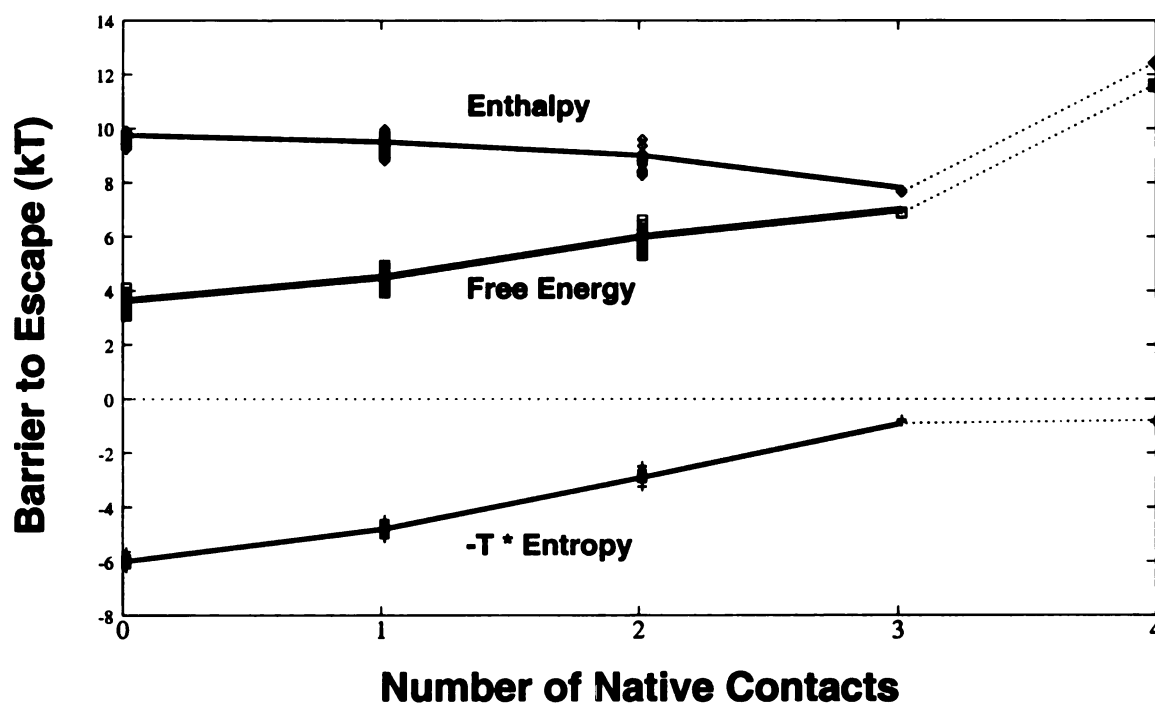
Figure 2.3: Size and components of total barriers to leaving conformations, as a function of depth in the landscape. (For sequence ABCAXCXCB.) [This shows that increase in leaving rate is due to a reduction in ways out. (The enthalpy is calculated as the Boltzmann weighted average of ΔH_{ij}, where ΔH_{ij} is the barrier height for jumping between a given conformation and all of the barriers to change to every other conformation.) The entropy component of this barrier measures the number of ways out, by indicating how many of the barriers in the transition state ensemble are actually likely to be taken by the folding chain. This number goes down with depth in the landscape. (The entropy is calculated by $\sum_i k_{ij} \ln k_{ij}$, where $k_{ij} = \exp(-\Delta H_{ij})$.)]

lifetimes. The number of exit routes goes down with the number of native contacts. Because of the funnel shape of the density of states, few conformational transitions are possible deep in the landscape.

Hence the folding kinetics results from a balance of two factors. The conformation-to-conformation transitions get faster deeper on the landscape because as the chain becomes more native-like, smaller structural rearrangements are needed for the chain

to become native. In this sense, folding is 'cooperative'. On the other hand, deeper on the landscape there are fewer routes to the native state. This entropic factor results in a slowing at the bottom of the landscape.

Even so, finding viable moves is not an especially slow process, even at the end. The probability of stepping to the native state is simply the probability of a particular microscopic transition. The only reason that these late stages are rate limiting is because the rates downhill from higher on the landscape are even faster. We refer to the behavior at the top of the landscape as 'entropic acceleration.' What we mean by 'acceleration' is just that there is a multiplicity of microscopic trajectories, hence a speed-up relative to a single trajectory. Here's the basis for entropic acceleration. Consider the folding of an unfolded chain. Suppose torsion angle j has an intrinsic rate k_0 of converting to its native $\phi\psi$ conformation. The whole chain can become native-like much faster than this intrinsic rate because there are N different torsion angles in the chain, all converting at a rate k_0, so the overall rate downhill for the chain is Nk_0. The overall flow downhill is faster than any one microscopic channel. At the bottom of the landscape, the multiplicity of routes is much smaller than the factor of N because the degrees of freedom become coupled.

Here's the difference between barrier-based kinetics, on the one hand, and entropic acceleration on the other hand. In barrier models, the smallest microscopic step defines the upper speed limit, kT/h. Barrier processes are slower than this intrinsic speed limit, approaching it only in the limit of high temperature, $T \to \infty$.

But in a funnel, the overall rate of folding is faster than the intrinsic microscopic conformation-to-conformation transition rate, because of the parallelization of microscopic trajectories. Such parallelization does not apply to classical processes that have a single dominant micropathway.

2.3.2 Eigenvectors: the degrees of freedom of the *ensemble*.

How should we describe the time-evolution of the populations of microconformations? We do this using the eigenvectors of the master equation. Each eigenvector is a vector of all the conformations, with specific weights. One eigenvector represents the equilibrium distribution of microstates. Its associated eigenvalue is zero, because that distribution is independent of time. All the other eigenvectors of the rate matrix represent deviations from the equilibrium distribution. Unlike the population vector, the components of an eigenvector can be negative. If an eigenvector component is positive, it means that conformation is more populated than in the equilibrium distribution. A negative value means the conformation is less populated. Because an eigenvector is only determined up to a scalar multiplier, the signs are arbitrary. What is significant is that in every eigenvector, some conformations have weights of opposite sign. For a system to be out of equilibrium, some conformations have to be over-represented and some under- represented.

The eigenvectors form a basis set for describing any population of microstates. The special property of the eigenvectors as a basis set is that they each have a simple

time dependence; their amplitudes all decay as single exponentials. The eigenvalue that corresponds to each eigenvector indicates the rate of decay: the inverse of the eigenvalue is the rate constant. So the relaxation process of a system can be viewed as the single exponential decay of the eigenvectors, until only the equilibrium eigenvector is left. Its amplitude does not decay with time because its eigenvalue is equal to zero.

An example

Here we give an example in a very simple system to illustrate the properties of eigenvectors. Consider a system that has only two states. We call them D and N. Any configuration of the system can be represented by indicating the population of each state: e.g. $.8D, .2N$. In vector notation we can say the standard population basis is $D = (1, 0)$ and $N = (0, 1)$. This means we can now represent the population, $p(t)$ as a vector, $p_{pop}(t) = (.8, .2)$. Next we generate a transition matrix for this system. Suppose D has a relative free energy $\Delta G = G_D - G_N > 0$ where G_N is the free energy of the native state. Now apply the Metropolis criterion for a transition. That is, $k(N \to D) = F$ and $k(D \to N) = Fe^{(-\Delta G/kT)}$, where F is the intrinsic rate of a barrierless transition (the front factor in Transition State Theory.) This leads to the transition matrix

$$\mathbf{K} = \begin{pmatrix} -F & Fe^{(-\Delta G/kT)} \\ F & -Fe^{(-\Delta G/kT)} \end{pmatrix}$$

The diagonal elements are the leaving rates, that is, the sum of all ways out. This can be found by the negative sum of all elements in the column of that diagonal, which

in this case is one element. To be more concrete, set $F = .1/(\text{unittime})$ and $\Delta G/kT$ to $ln.5$, and F to $1/1.2$:

$$\mathbf{K} = \begin{pmatrix} -.1 & .05 \\ .1 & -.05 \end{pmatrix}$$

The Master Equation represented by the matrix K says that .1 of D goes to N per unit time, and .05 of N goes to D per unit time. We now find the eigenvectors and eigenvalues of this matrix:

$$\mathbf{K} = \mathbf{M}^{-1}\mathbf{L}\mathbf{M} = \mathbf{M}^{-1} \begin{pmatrix} .15 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -1 & 2 \end{pmatrix}.$$

The diagonal elements of the matrix $\mathbf{L}$ are the eigenvalues, and the columns of the matrix $\mathbf{M}$ are their associated eigenvectors:

$$\lambda_1 = .15$$

$$\lambda_0 = 0$$

and

$$v_1 = \begin{pmatrix} 1D \\ -1N \end{pmatrix}, v_0 = \begin{pmatrix} 1D \\ 2N \end{pmatrix}.$$

In this notation we have labeled the usual conformation vectors D and N. As expected one eigenvalue is equal to zero. Its associated eigenvector (normalized) gives the equilibrium distribution: $(1/3D, 2/3N)$. Also, the second eigenvalue reproduces the standard kinetic result for the relaxation rate constant: $\lambda_1 = .15 = k_{\text{obs}} = k_f + k_u = .1 + .05$. So all of the conventional kinetic information is contained in the eigenvectors in and eigenvalues.

But the eigenvectors also give additional information. In this case, the non-equilibrium eigenvector reveals that the two-states are not kinetically independent. The eigenvector $v_1 = (1, -1)$ shows that any over-representation of D must be exactly compensated for by a reduction in N. Taken together, the eigenvectors reveal the underlying kinetics of the system. One eigenvector (with an eigenvalue equal to zero) gives the equilibrium distribution of the system. The other eigenvector says that the two states are linked, and that one decays into the other with the expected rate constant. In contrast, if we use the untransformed conformations as the basis set, we would have two independent quantities, D and N, and we need an extra equation $(D + N = 1)$ to relate them. But this constraint relation is already contained in the eigenvectors. In the eigenvector framework, the system has only one degree of freedom (see Figure 2.4). Now we apply a similar analysis to our more complex folding model having 740 conformations.

2.3.3   Ensemble vs. conformational degrees of freedom

The eigenvectors can provide information on the shape of the folding funnel. At the top of the funnel, most of the degrees of freedom (bond angles) are independent of each other; at the bottom (the native state), no conformational freedom remains. Folding is a process by which the degrees of freedom become coupled to each other. Said differently, the question of why protein folding kinetics is 2- state can be recast as the question of why only one eigenvector dominates the rate of emergence of the
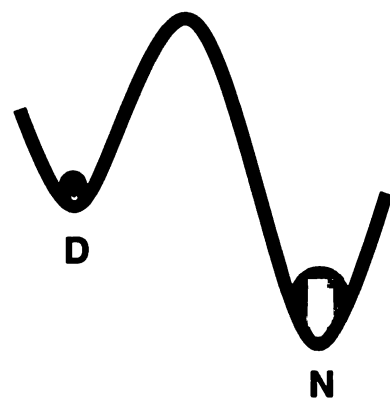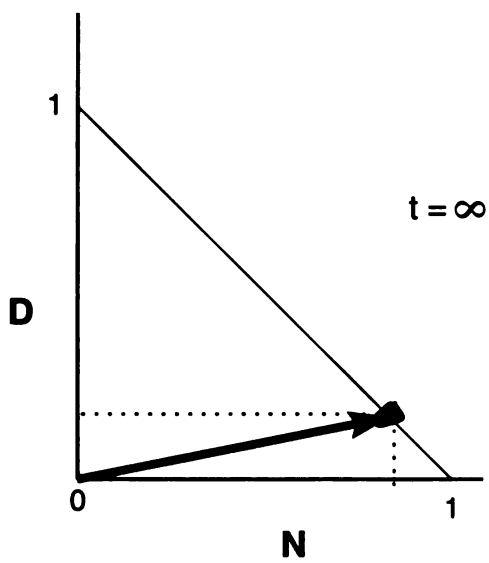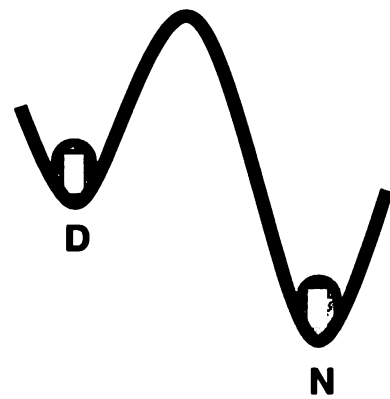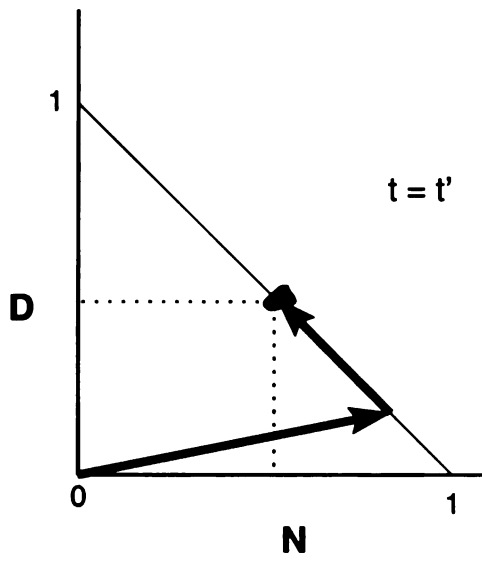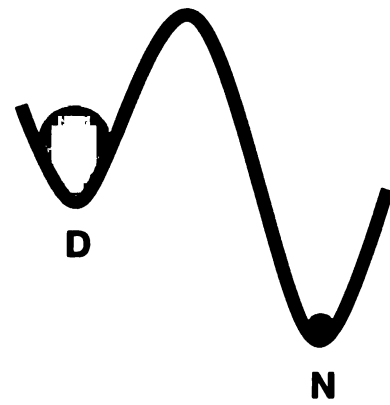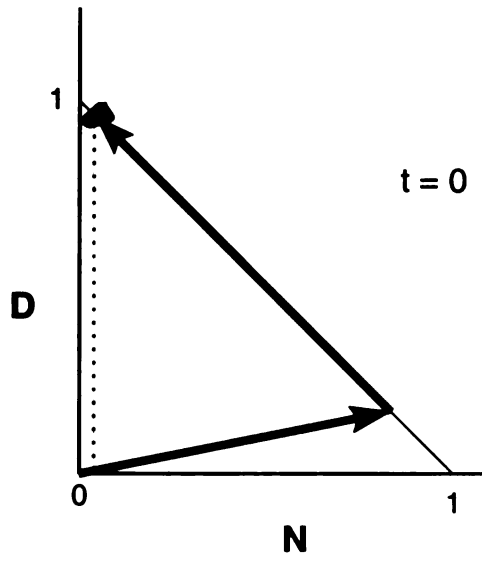
native state. Since each eigenvector represents a degree of freedom of the ensemble, apparent two-state folding is single- eigenvector folding.

To address this question, we develop a strategy akin to the Boltzmann *H-theorem* [29]. We determine the coupling of the degrees of freedom of the out-of- equilibrium ensemble by an entropy-like quantity:

$$S_{\text{kin}}(t) = \sum_i w_i(t) \ln w_i(t),$$

where $w_i(t)$ is the weight of eigenvector i at time t. We determine $w_i(t)$ by

$$w_i(t) = \frac{\|v_i(t)\|}{\sum_i \|v_i(t)\|},$$

where $\|v_i(t)\|$ is the component of the ensemble in the direction of eigenvector i at time t. These magnitudes change with time, decaying, according to their eigenvalues, from the initial values, which are determined by the initial ensemble. These weights measure the displacement of the ensemble from equilibrium in various independent directions. We label this S_{kin} because it reflects the number of degrees of freedom of the ensemble that are significantly away from equilibrium. This is quite different from the entropy in conformational space ($S_{\text{conf}} = \sum_i p_i \ln p_i$,) as can be seen in figure 2.5. (We define the entropy of a non- equilibrium ensemble operationally by S_{conf} above; the time dependent p_i are not Boltzmann weights either.) Physically, S_{kin} gives a measure of the number of "characteristic" directions in which ensemble differs from equilibrium.

Figure 2.5 shows how degrees of freedom are lost from the ensemble with time. At
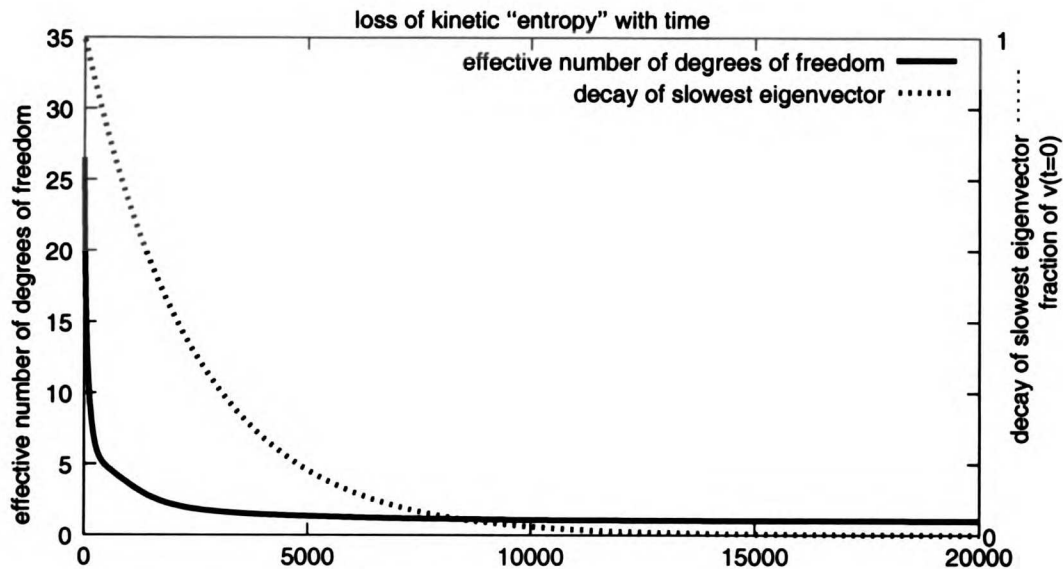
Figure 2.5: Decay of the number of degrees of freedom relative to the decay of the slowest degree of freedom.

early times, the ensemble is displaced from its native equilibrium in many directions (large S_{kin}.) As the system approaches equilibrium, only the directions with slow relaxation rates (those with small eigenvalues) will remain significant. Most remarkably, the degrees of freedom become coupled much faster than the relaxation time of the ensemble to the native state. At infinite time, the ensemble vector coincides with the equilibrium ensemble vector, so $S_{kin} = \ln 1 = 0$.

The fastest decaying eigenvectors mostly contain deviations from equilibrium among the weights of conformations with the fastest leaving rates. The slower eigenvectors account for deviations from equilibrium in the weights of lower energy conformations, which have slower leaving rates (see Figure 2.6). We find that there can be large changes in the ensemble very quickly. This is consistent with kinetic studies
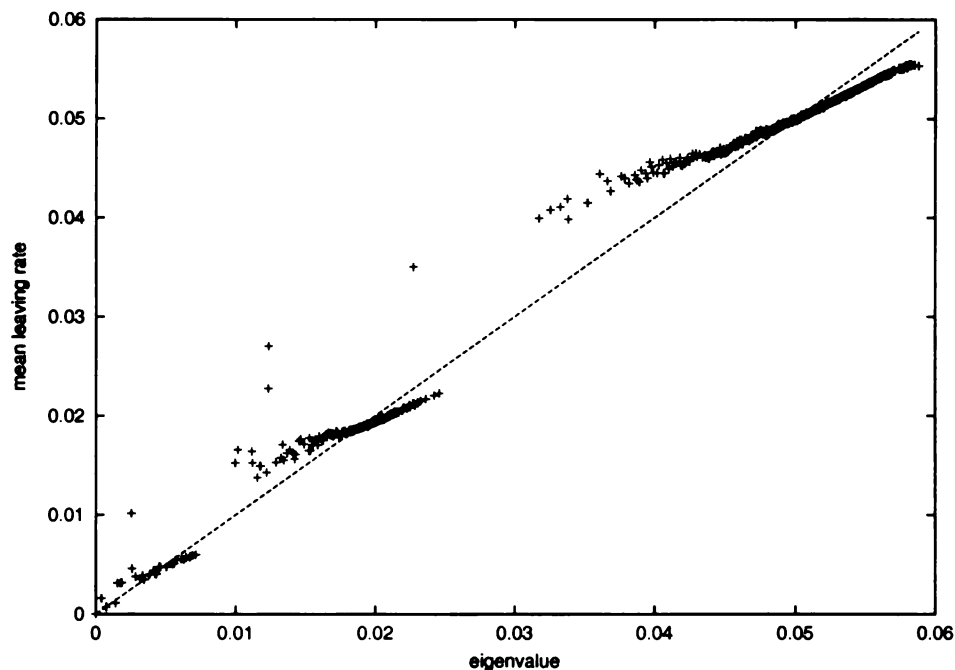
Figure 2.6: Relationship between the eigenvalue and the mean leaving rate of the distribution of conformations represented by their associated eigenvectors. Eigenvectors with faster eigenvalues govern conformational distributions with higher mean leaving rates.

that show much structural change in the dead time of re-folding experiments [30].

2.3.4 Why are rate constants *constant*?

Our model predicts chevron behavior, a key experimental result that has been used as evidence for 2-state behavior. What is the physical basis for it, in this model? We first address a more elementary issue. In 2-state models, rate coefficients such as $k_{forward}$ or $k_{reverse}$ are independent of time. In a mass- action model, a rate coefficient describes the ratio of the rate of approach towards equilibrium to the deviation from

equilibrium:

$$\frac{\frac{d}{dt}p(t)}{p(\infty) - p(t)} = k_{obs}.$$

For a single exponential process this ratio is constant: the exponential time dependence cancels with its derivative. The conventional physical explanation for single exponential kinetics is that there is a free energy barrier high enough that it is possible for the configurations on either side of it to come to internal equilibrium faster than they interconvert. But pre- equilibration is not required to get single- exponential kinetics.

Any distribution that deviates from equilibrium in such a way that the relaxation back towards equilibrium goes at the same rate for conformations will give single-exponential kinetics. Each of the eigenvectors represents just such a deviation from equilibrium. Here we demonstrate that is the case. The "pre-equilibrium" distribution is not the only distribution that that will give rise to steady-state (i.e. single exponential) kinetics. Consider a distribution that is perturbed from equilibrium only along the direction of the slowest non-zero eigenvector. First note that, since multiplying my the rate matrix is equivalent to taking the derivative:

$$\frac{\frac{d}{dt}p(t)}{p(\infty) - p(t)} = \frac{\mathbf{K}p(t)}{p(\infty) - p(t)}.$$

We can deconvolve $p(t)$ into the sum of the equilibrium distribution and the slowest eigenvector multiplied by a time dependent factor α. We define v_0 as the equilibrium

eigenvector, and v_1 as the eigenvector corresponding the first non-zero eigenvalue.:

$$p(t) = v_0 + \alpha(t)v_1 = p(\infty) + \alpha(t)v_1.$$

Substituting this in yields

$$\frac{\frac{d}{dt}p(t)}{p(\infty) - p(t)} = \frac{\mathbf{K}v_0 + \mathbf{K}\alpha(t)v_1}{-\alpha(t)v_1.}$$

But this can be simplified because $\mathbf{K}v_0 = 0$. Also, $\mathbf{K}$ is a linear operator, so the $\alpha(t)$ in the numerator can be moved through and hence canceled with itself in the numerator, removing the time dependence:

$$\frac{\frac{d}{dt}p(t)}{p(\infty) - p(t)} \frac{\mathbf{K}v_1}{-v_1}$$

Finally, $\mathbf{K}v_1 = \lambda_1$, and, rewriting the left hand side as k_{obs}:

$$k_{obs} = -\lambda_1.$$

So k_{obs} can be independent of time even when $p(t)$ is not. The folding rate in this case, with perturbations from equilibrium only in the direction of the slowest eigenvector, is given by the slowest eigenvalue. Even though there is not a large barrier allowing for true pre-equilibration, exact single-exponential kinetics are achieved. Pre-equilibration is not required to give a single- exponential folding.

In the slowest eigenvector, the lowest energy conformation will be of opposite sign compared to all the other conformations (see Figure 2.7. This means that the folding rate is determined by the average of folding rate of all non-native conformations.
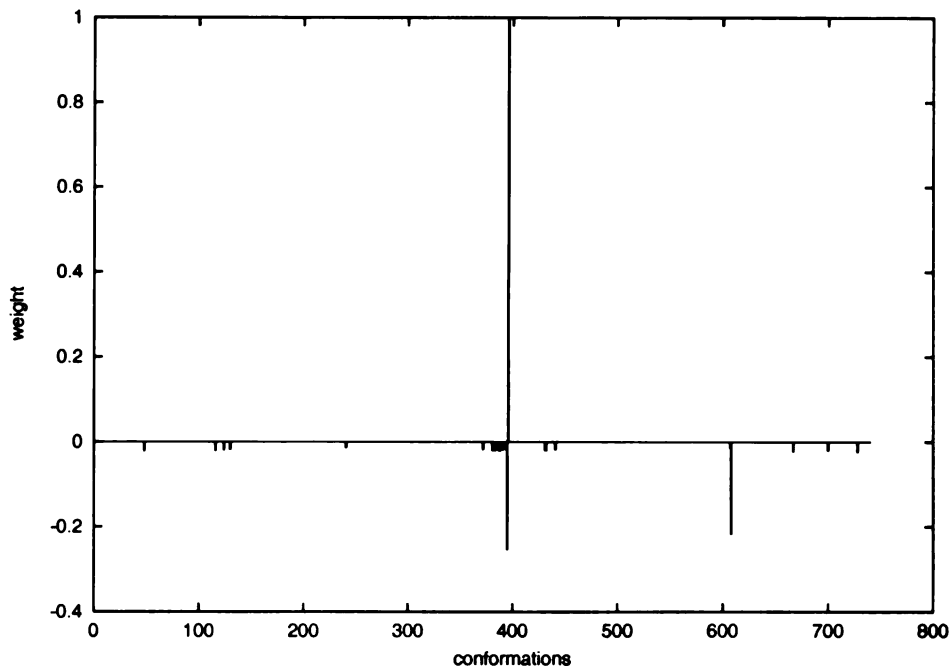
Figure 2.7: The slowest non-zero eigenvector. The large positive conformation in the lowest energy conformation. All other conformations have opposite in sign to this one, and hence this vector represents the interconversions between the native configurations and all others.

That is:

$$k_f = \frac{\sum_{i \neq N} p_i k_{i \to N}}{\sum_{i \neq N} p_i},$$

where p_i is the weight of the ith conformation in the vector v_1 These p_i's are the *deviations* from the equilibrium values of the weights. This differs from the actual time dependent weights of these conformations. Single exponential kinetics occurs whenever these deviations all change by the same amount in proportion to their size, irrespective of that size. k_{obs} is independent of time, because the direction of the deviation is independent of time, even though the magnitude of the deviation is not. This means that the distribution of unfolded conformations changes with time. It re-

laxes to the equilibrium distribution. The unfolding rate is simply $k_u = -\sum_{i \neq N} k_{N \to i}$, independent of time, and of the instantaneous distribution of conformations.

We are now in a position to rationalize the shapes of the chevron plots. The observed relaxation rate is the sum of an unfolding rate and a folding rate. In our model, the individual forward folding rate terms, $k_{i \to N}$, are independent of contact strength. They are based only on the configurational difference between the native configuration and the ith configuration. What changes the folding rate are the p_i terms: the weight of each conformation in the refolding ensemble. The more the refolding ensemble is biased towards conformations with greater geometric similarity to the native state, the faster folding will be. The conformation-to- conformation barriers get smaller as structures become more native-like.

The apparent transition state barrier is an average of all the $k_{i \to N}$'s. How does this apparent barrier change with denaturant or temperature? As external conditions become increasingly native-like, the denatured ensemble becomes increasingly native-like. Since the starting population is more native- like, folding is faster. As conditions change, the Boltzmann distribution of denatured conformations changes. The apparent transition state is a kinetic distribution that does not shift identically with the Boltzmann distribution. Figure 2.7 shows a typical dominant eigenvector (apparent transition state). It largely tracks the changes in the Boltzmann distribution in the denatured state, except at the lowest energy non-native conformations, which are slightly under-represented, even though they dominate the refolding ensemble.

2.3.5 Chevron Plots

Our model predicts a "chevron" dependence on the strength of the contact energy (Figure 2.8). The logarithm of the folding rates increases linearly as the contact energy gets stronger, and the unfolding rate increases linearly with contact energy above the denaturation midpoint. Under strongly native conditions, our chevron plots roll-over, as in experiments [31, 8, 32].

Mass-action models assume that the two relevant states D and N are fixed, not dependent on denaturant or temperature. Changing denaturant simply changes the driving force for the D molecules to bury exposed hydrophobic surface area [33]. But in our model, the ensembles do change with contact strength. In our model, the denatured ensemble changes more with denaturant than the transition state. Figure 2.8 shows the dependence of the dominant relaxation rate on contact strength, for the first native state depicted in Figure 2.11. (Other sequences produce similar dependencies.)

Chevron plots and two-state folding

Observing chevron behavior is not proof that a system has two fixed states. For a two-state system, the ratio the folding and unfolding rates must equal the equilibrium constant, and their sum must equal the observed relaxation rate. That is, we have:

$$k_f = \frac{K_{eq}k_{obs}}{1 + K_{eq}},$$

and

$$k_u = \frac{k_{\text{obs}}}{1 + K_{\text{eq}}}.$$

But if a system is not 2- state, these relations can be satisfied, even though they no longer uniquely specify k_u and k_f.

Curved chevron plots do not necessarily mean intermediates.

Curves in chevron plots can be consistent with two- state equilibrium data[34]. In fact, any chevron plot can be consistent with equilibrium data. True two-state folding means that $k_f/k_u = K_{eq}$, while $k_{\text{obs}} = k_f + k_u$. But it is always possible to find a k_f and k_u such that this is true, because this is simply a case of two equations and two unknowns: $k_f = k_{\text{obs}}K_{eq}/(1 + K_{eq}), k_u = k_{\text{obs}}/(1 + K_{eq})$. Thus, observing a chevron plot is not proof of two-state folding. It has been suggested that roll- over in chevron plots indicates the presence of an intermediate on the folding pathway. The rationale is that if there was no intermediate, the transition state barrier would continue to shrink with increasingly native conditions, and roll-over would not appear[33, 35, 10, 36].

Roll-over occurs, too, in our simulations. It results from a saturation of the non-native conformations that flow to the native structure. Ultimately, the whole ensemble of non-native states becomes these 'supply conformations', and no further population shifts are possible (see Figure 2.8).

Thus we find that time-scale separation is an inherent property of some funnel-
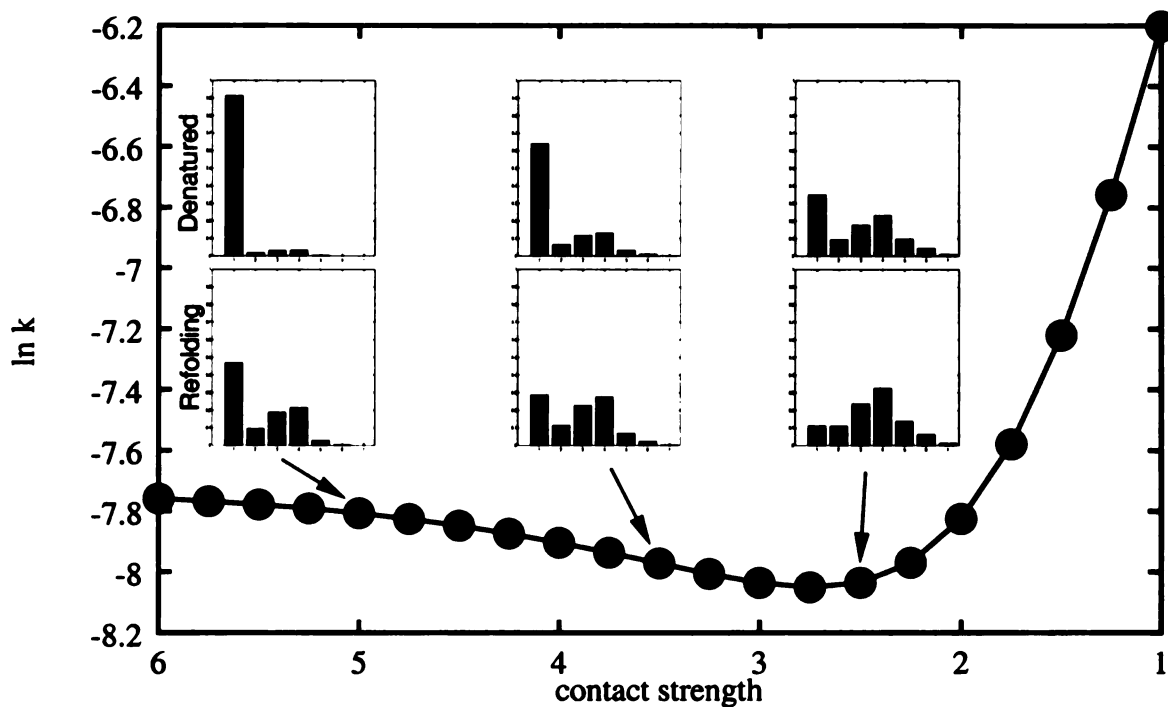
Figure 2.8: Folding rate vs. contact strength. This figure shows the dependence of the rate of equilibration on the strength of contacts (in units of kT.) This is analogous to "chevron" plots of relaxation rate vs. denaturant concentration. The insets show the nativeness of the denatured and refolding ensembles at several temperatures.

shaped landscapes, and that the model results are consistent with a key property observed in experiments.

2.3.6 The Search for a Free Energy Barrier

Figure 2.12 shows the free energy profiles as a function of the number of native contacts, for different values of the contact strength, ε/kT. When ε/kT is small, when only 20% of the ensemble is in the native conformation at equilibrium, there is a barrier to folding. But under more native- like conditions, the figure shows that the barrier disappears: folding is thermodynamically downhill. We find single-exponential kinetics in both cases.

2.3.7 When Individual Trajectories Will Not Capture the Kinetics

For proteins that fold when true barriers are present—energetic or entropic, then studying one or more individual trajectories in molecular dynamics or Monte Carlo modeling [37, 38, 39] can reveal the bottleneck conformations. On the other hand, if 2- state protein folding is governed by entropic acceleration of the type described here, then analyzing trajectories will not give much information about the rates. In entropic acceleration, it is the *number of trajectories*, and the degree of parallelism of the micropaths, and the reduction of that parallelism down the energy landscape, that determines the kinetics, not barriers along any particular micropathway. Of course, it is possible that both barrier processes and entropic acceleration can contribute to the folding kinetics of any real protein.
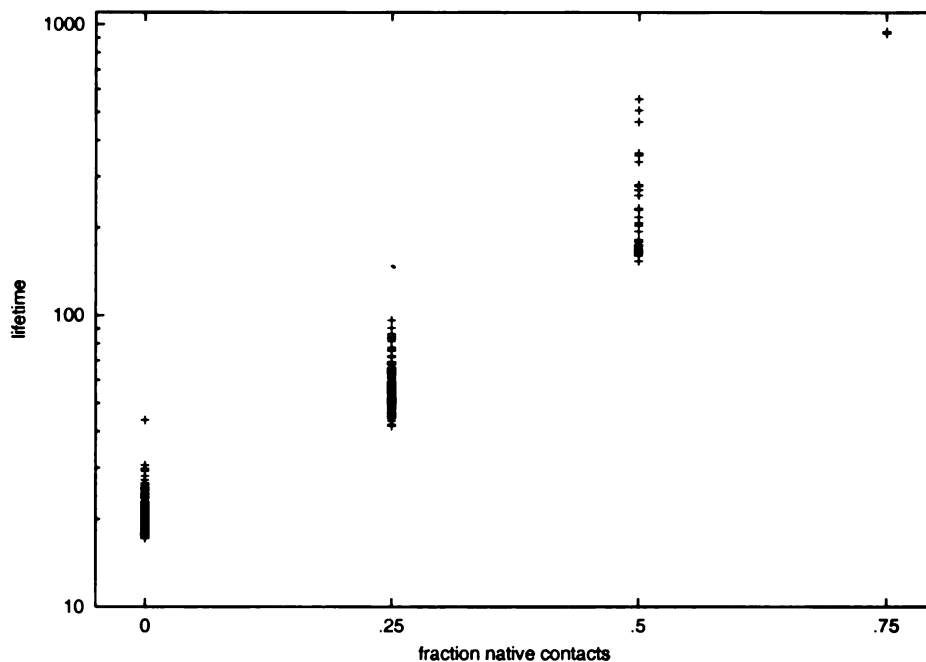
Figure 2.9: Conformations that are deeper in the landscape (i.e. those with more native contacts) have exponentially longer lifetimes.

Our modeling shows that the deeper a conformation is on the energy landscape (more native-like), the longer its lifetime. (See Figure 2.9.) It is interesting that in the longest MD trajectory to date of a protein in explicit water [38], the most-native-like structure observed was also the longest-lived.

Comparison with Transition State Theory

Transition-State theory or Kramers theory parse the rate constant into two factors:

$$k = fp(TS),$$
(2.1)

where f, the front factor (usually kT/h) is the rate of going forward, and $p(TS)$

is the very small probability that the system has reached this high-energy state TS, $p(TS) = exp(-\Delta G/k_B T)$, where ΔG is the free energy difference to the transition state from the reactant. All the populations relevant to the kinetics are determined by the thermodynamic premise, that the transition state acts like an equilibrium state.

Our simulations give a different interpretation. Ours is a kinetic model that does not rest on a thermodynamic premise. Single- exponential kinetics arises here, even in the absence of a pre- equilibration assumption. The ensemble of non-native conformations changes continuously with time, as they become increasingly native- like. Neither the denatured state nor the apparent transition state are fixed unchanging distributions, as would be associated with the symbols in mass-action model equations.

There are three ensembles that are often assumed to be the identical in mass-action modelling of protein folding kinetics: (1) The Boltzmann ensemble of non-native conformations under the initial unfolding conditions, (2) The Boltzmann ensemble of non-native conformations under the final folding conditions. Both of these are usually taken to be essentially random coil distributions. (3) The distribution of conformations from which folding takes place, which is not an equilibrium ensemble. In short, mass-action models envision macrostates that are specific invariant distributions of microstates, and folding is seen as a flowing from one such macrostate to another. Our model shows, instead, that the whole folding process is simply the continuous time-evolution of a single ensemble, as the non-native microstates become native.

2.3.8 Comparison to the nucleation-condensation model

The transition state of a protein has been called a 'nucleus', in analogy with nucleation processes [4]. In this interpretation, the high energy barrier, by definition, corresponds to the least-probable conformation of the molecule. Once a protein has a nucleus of sufficient size (reaches the top of the barrier), folding proceeds rapidly (condensation).

But that description is in terms of the reaction-coordinate diagram, not the microscopics as represented by the energy landscape. We find that the rate limiting chain conformations are not necessarily rare. Rather, in the funnel model, these rate-limiting conformations can be the most populated conformations in the unfolded ensemble. On a relative basis, the populations can be small numbers, but since there are 10^{15} - 10^{50} possible conformations for a typical protein, these gateway conformations can be the largest single conformations. Since they are less stable than the ensemble of all other possible conformations, they would constitute a bottleneck on the folding landscape. But their relative improbability is not necessary to generate single exponential kinetics. And viewing them as points of maximum free energy along a trajectory is misleading. These may be points of high free energy on a reaction- coordinate diagram, if they are represented as a Transition State (TS) between all other non-native conformations (D) and the fully folded conformation (N). But along a single trajectory they can have low free energy, second only to the native conformation.

2.4 Conclusions

Current explanations of 2-state protein folding kinetics are based on the idea that there is a barrier—energetic or entropic—implying that there is some small set of special transition-state conformations that are in low population and that govern the rate of folding. Much effort has been invested in searching for these transition state conformations. Here we develop a rigorous way to identify the rate-limiting steps in processes like protein folding that involve multiple microscopic processes that can happen in parallel. It gives a different explanation for single-exponential kinetics, and for chevron plots, the main experimental fingerprints of 2-state processes. Two-state kinetics is explained instead in terms of entropic acceleration. Folding happens faster than intrinsic bond rotation rates because there are so many different bonds in the chain that are free to change at the same time. Folding happens by parallel processing. The multiplicity of routes means very fast flow at the top of energy landscapes, but the multiplicity becomes smaller at the bottom. This gives a separation of time scales that is dominated by a single- exponential. One implication is that studies of single trajectories in models may not be sufficient to describe the kinetics.
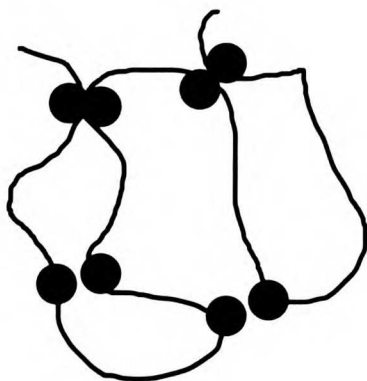
2.5 Methods

Thermodynamics: Gō potential. For structural and thermodynamic properties we use the Gō model, in which proteins are modeled as two dimensional chains on
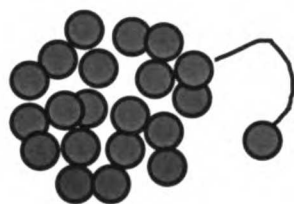
A.
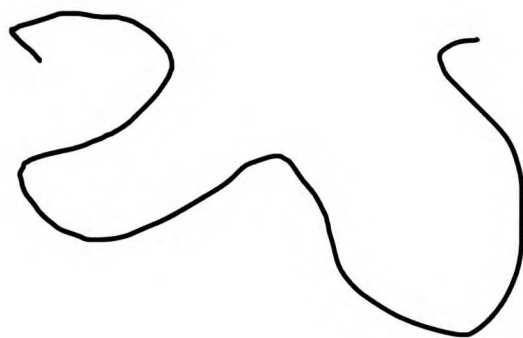
High probability of moving to another conformation.
High probability of making a contact..
Low probability of jumping to native.

B.

Lower probability of changing conformation.
Lower probability per unit time of making a new contact.

C.

Lowest probability of changing conformation.
Highest probability that the change will be to native.

Figure 2.10: Summary: (**A**) Open conformations have a high probability of changing to new conformations, and high probabilities of making contacts, because there are many ways they can. (**B**) As the chain makes contacts, its motion becomes restricted. This means that its probability of changing conformation is reduced, as many random kicks would break favorable contacts, so it takes it longer to get pushed in an acceptable direction. (**C**) As the chain gets very close to its native structure almost all changes in conformation would take it uphill in energy. The rate of change at this point reduces to the probability of the last move necessary to make it the native structure. The folding rate will depend on the occupancy of these nearly folded states in the refolding ensemble. If they are rare this reduces to the conventional transition state picture. But it is not necessary to assume that they are rare. Even if they are favored relative to more open states, one should see simple kinetics.
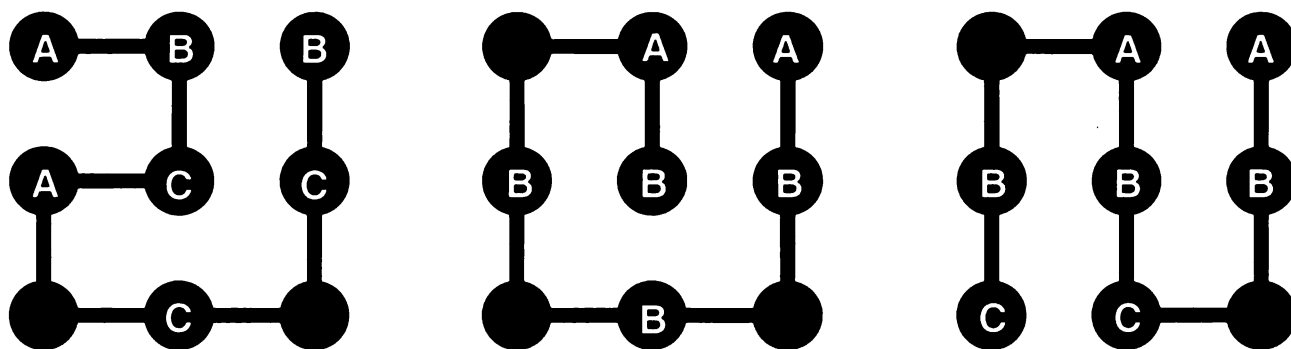
Figure 2.11: The three maximally compact configurations of chains with nine beads. The letters on the beads indicate "residue types." In the Gō potential only residues of the same type have favorable energy of interaction. (The blank beads make no contacts in the native structure, and so have zero contact energy with all residue types, including other blank beads.)

a square lattice [17]. Beads that occupy neighboring sites on the lattice, and are not connected by links of the chain, are in contact. Contacts present in the native conformation have a favorable contact energy, $-\varepsilon$. All other contact interactions have zero energy. With $\varepsilon > 3kT$ this potential generates "good" thermodynamic funnels for all of the maximally compact conformations of a chain with nine beads. (Figure 2.11 shows these conformations) Under these conditions, the Gō potential generates good thermodynamic funnels because the entropy lost upon making a native contact is always more than compensated by the favorable contact energy. Figure 2.12 shows the free energy profiles for the energy landscape at different contact strengths.

To solve the kinetics, we use a Master Equation. The total conformational space is 740 lattice conformers. The rate of transition between conformation i and j is determined by a set of elementary moves. Though there has been controversy in
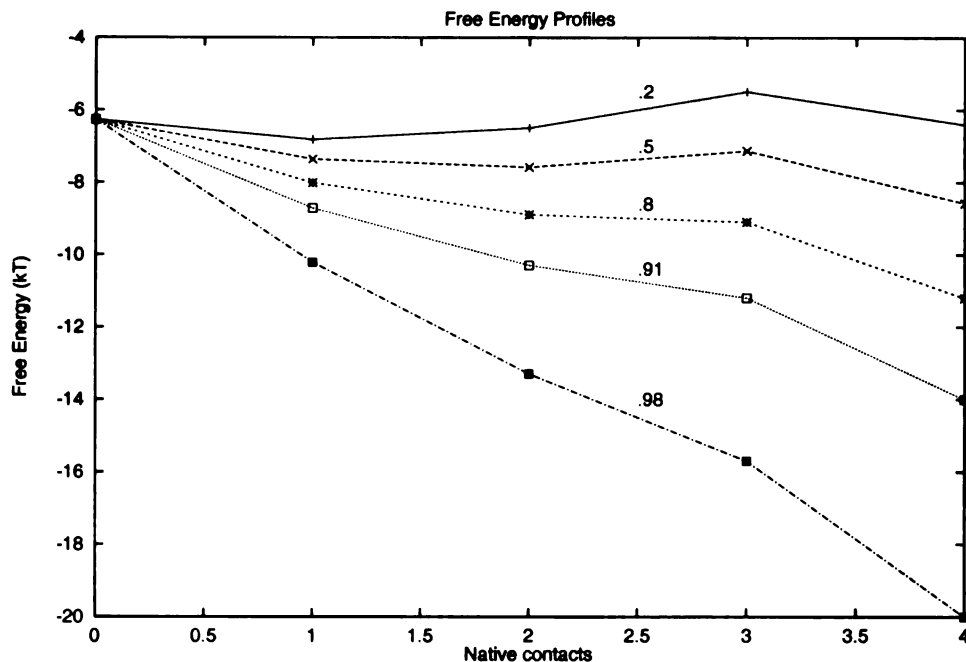
Figure 2.12: The free energy profiles at various contact strengths. The labels next to each curve give the fraction the ensemble that is in the native conformation at each contact strength. At unstable contact strengths the profiles show a barrier. Above about 80% native stability this barrier flattens out. Folding, however, still obeys a single exponential.

applying an equilibrium methodology to a kinetic process, and care must be taken to ensure the validity of choices made in modelling microscopic kinetic events, [40, 41, 42] these methods have proven to be useful. These transition probabilities taken together determine a set of differential equations, collectively defining the Master Equation. Master Equations are the most rigorous method for studying the relaxation of a system [43] and have been increasingly applied to the protein folding problem [44, 45, 46, 2]. The solution to the Master Equation gives the distribution of probabilities in the whole ensemble conformations, as a function of time, given an initial distribution.

We find the solution numerically, in terms of the eigenvectors and eigenvalues of the master equation matrix [28, 27]. This gives the full population kinetics, which is equivalent to what is measured in refolding experiments [4].

Kinetics: transition probabilities. We model the barrier for the transition from one conformation to another as being due to: (1) bond rotation barriers and (2) motion through the solvent. For a protein to change to a new conformation it must gain enough free energy to rotate bonds and displace the solvent. In addition, the chain needs to gain enough free energy to disrupt any favorable contacts that are broken in the transition. We assume that all changes are independent: the probability of a move is the product of the probabilities of the individual changes necessary to make it happen. We calculate the individual barriers in the following way: Each angle that differs between two structures requires the surmounting of a rotation energy barrier of height h. So the total barrier due to differences in angles is

$$\Delta H_{ij}^{\ddagger \text{(angles)}} = h \sum_{n=1}^{N-2} \delta(\theta_{in}, \theta_{jn}) = h \Delta \Theta_{ij},$$

where $\delta(\theta_{in}, \theta_{jn})$ is zero if $(\theta_{in} = \theta_{jn})$ and 1 otherwise.

Similarly, the total barrier to a given displacement of bead positions is the sum of the probabilities of each individual bead diffusing to its new position:

$$\Delta H_{ij}^{\ddagger \text{(positions)}} = \gamma \sum_{n=1}^{N} (\|\vec{r}_{in} - \vec{r}_{jn}\|^2) = \gamma \Delta R_{ij}^2$$

where γ characterizes the viscosity of the solvent. The $\vec{r}_{in}$ are measured from the center of mass of each conformation, after optimal alignment on the lattice.

The total barrier height for each conformation-to-conformation transition is the sum of all these changes:

$$\Delta H_{ij}^{\ddagger} = \begin{cases} h\Delta\Theta_{ij} + \gamma\Delta R_{ij}^2 + \varepsilon\Delta n_{ij} & \text{if } \varepsilon\Delta n_{ij} > 0 \\ h\Delta\Theta_{ij} + \gamma\Delta R_{ij}^2 & \text{otherwise.} \end{cases} \tag{2.2}$$

This barrier height can be converted to a first order rate constant for each microscopic transition using transition state theory:

$$k_{ij} = \frac{1}{\tau D} \exp(-\Delta G_{ij}^{\ddagger}/kT).$$

This satisfies detailed balance because $k_{ij}/k_{ji} = \exp(-\varepsilon\Delta n_{ji}/kT)$. The front factor, τD, contains two terms. One is the absolute rate factor that serves to set the unit of time (τ). The other, D, is a diffusion constant, a normalization factor that is necessary because the rate of a transition is not given simply by the the height of a barrier, but also by the probability that the degrees of freedom are excited in the direction that would lead to a particular jump. This reduces the overall rates. However, this term must be the same for all conformations by microscopic reversibility, so we do not need to know its actual value. We simply employ an arbitrary front factor $F = \tau D$, which sets the units of time. In practice we set F equal to the number of conformations of the chain. This insures that the sum of the rates out of any conformation is less than one, so they can be represented as probabilities per unit time.

We chose $h = .5kT$ and $\gamma = .1kT$, which gives relative transition probabilities similar to those in traditional lattice Monte Carlo studies of folding. Traditional

Monte Carlo move sets contain these barriers implicitly. Here we have made them explicit functions of two parameters. This facilitates our calculation of transition probabilities. Our goal is not to define the most accurate move set, but to explore the consequences of a reasonable one.

Setting up the Master Equation. The master equation is a set of differential equations, one for the rate of change of each conformation. It can be written in matrix form:

$$\frac{d}{dt}\begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ \vdots \\ p_N \\ \vdots \\ p_\Omega \end{pmatrix} = \begin{bmatrix} k_{11} & k_{12} & k_{13} & \cdots & k_{1\Omega} \\ k_{21} & k_{22} & k_{23} & & k_{2\Omega} \\ k_{31} & k_{32} & k_{33} & & k_{3\Omega} \\ \vdots & \vdots & \vdots & & \vdots \\ k_{N1} & k_{N2} & k_{N3} & & k_{N\Omega} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ k_{\Omega 1} & k_{\Omega 2} & k_{\Omega 3} & \cdots & k_{\Omega\Omega} \end{bmatrix} \begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ \vdots \\ p_N \\ \vdots \\ p_\Omega \end{pmatrix}$$

The diagonal elements of this matrix are given by:

$$k_{ii} = -\sum_{j \neq i} k_{ji}, \tag{2.3}$$

which insures that probability is conserved. Physically, k_{ii} is the rate at which conformation i would decline if no other conformations were converting into it. These *leaving rates* play a vital role in determining the overall kinetics, as will be shown below. A graphical representation of the entire equation is shown in Figure 2.13.

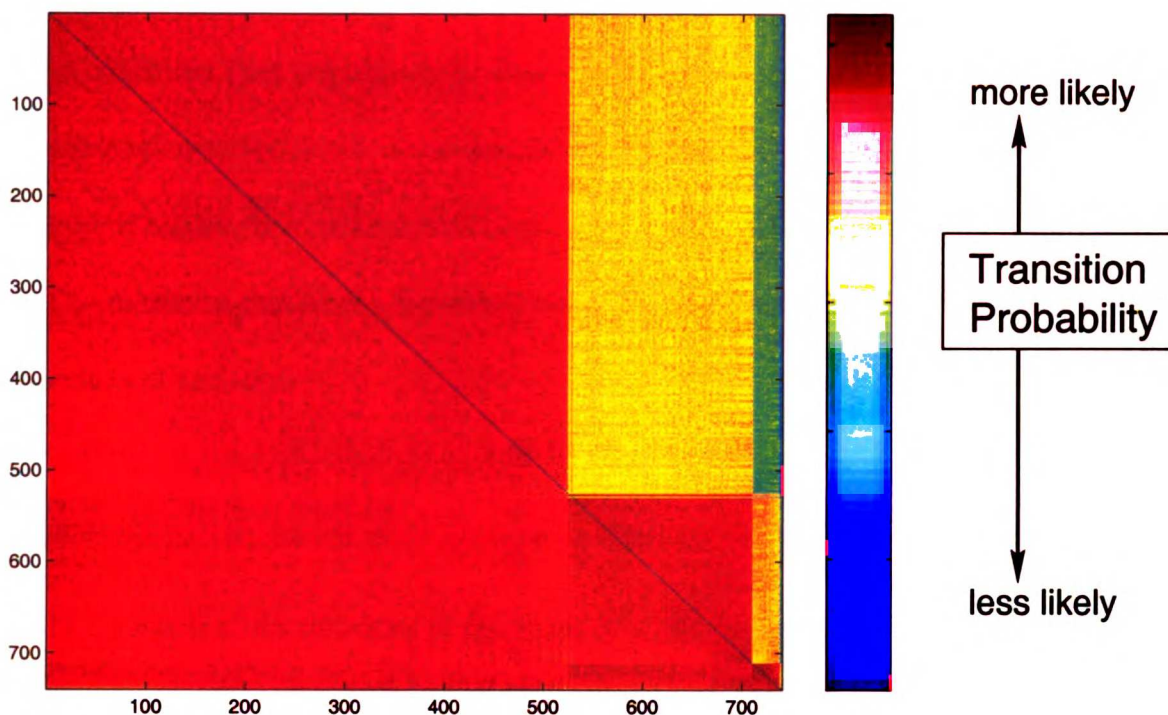Equation 2.5 can be rewritten as a differential equation:

Figure 2.13: This figure shows the entire master equation by color coding interconformational transition rates according to the log of their magnitude. The conformations are ordered by leaving rate, with the fastest at zero, and native at 740. This correlates with number of contacts. Conformations high on the landscape, with few contacts, have many ways out, as indicated by the their columns being all red. As the chain moves down in energy, a larger and larger fraction of transitions require breaking favorable contacts. This can be seen by the yellow and blue regions in the columns of conformations with more contacts.

$$\frac{d}{dt}\mathbf{p} = \mathbf{K}\mathbf{p}. \tag{2.4}$$

The matrix $\mathbf{K}$ is equivalent to the $\mathbf{T} - \mathbf{I}$, where T is the Transition Matrix defined in earlier work [40], and I is the identity matrix. The difference is simply whether multiplying by the matrix gives the change in population ($\mathbf{K}$), or the new population ($\mathbf{T}$).

The general solution to this system of differential equations, together with an initial condition (the population at time $t = 0$), gives the fraction of the population in each conformation at all times thereafter. To find the general solution we use the method of eigenvalues and eigenvectors.

The master equation can be solved using the eigenvalue- eigenvector method because it is of the form

$$\frac{d}{dt}\mathbf{p} = \mathbf{Kp}. \tag{2.5}$$

This equation has solutions of the form $e^{\lambda t}\mathbf{v}$. For every $\lambda, \mathbf{v}$ pair that satisfies:

$$\mathbf{Kv} = \lambda\mathbf{v}.$$

The solution of the system in equation 2.5 can be written as a sum of the eigenvectors of the master equation, multiplied by $e^{\lambda t}$ where λ is the eigenvalue associated with a particular eigenvector. The initial amplitude, or weight, of each eigenvector is determined by the initial conditions. We found the eigenvalues, eigenvectors, and initial weights for our master equations numerically using QL decomposition after Householder reduction to tridiagonal form [47]. We use algorithms for symmetric matrices on a reduced form of our master equation. The reduction is performed by multiplying each off-diagonal element by $e^{-\Delta E_{ij}/2}$. This gives a symmetric matrix. (The only asymmetric aspect of the matrix is the condition for detailed balance.) This reduced matrix has the same eigenvalues as the original, and its eigenvectors

can be converted back to those of the original matrix by multiplying each element by

$e^{-\Delta E_i}$ [28].

The amplitude of each eigenvector at $t = 0$ is determined by the initial probability distribution. We initiate folding with each conformation having equal probability at $t = 0$. This initial condition represents the distribution under strong denaturing conditions. At equilibrium in an ideal denaturant, native contacts should have no favorable energy of interaction, and all conformations have the same Boltzmann weight.

To determine how single-exponential simplicity can emerge from a system of differential equations containing $740^2 (= 547,600)$ rate constants, we examine the properties of the eigenvectors and eigenvalues, as well as the underlying energy landscape, as represented by the master equation itself.

We found that the fraction of the ensemble in the native conformation as a function of time is dominated by a single exponential. This is true for the three lattice conformations we tested, under a wide range of kinetic and thermodynamic parameters. Figure 2.14 shows a typical folding curve. (The results are similar under a wide range of conditions, so for simplicity we present most results for just one conformation under one set of parameters.) The folding curve is generated by summing the component of all eigenvectors that lie in the direction of the lowest energy (native) conformation. The folding curves are dominated by the slowest eigenvalue, with a small burst phase determined by the next faster eigenvalue. After a time interval that
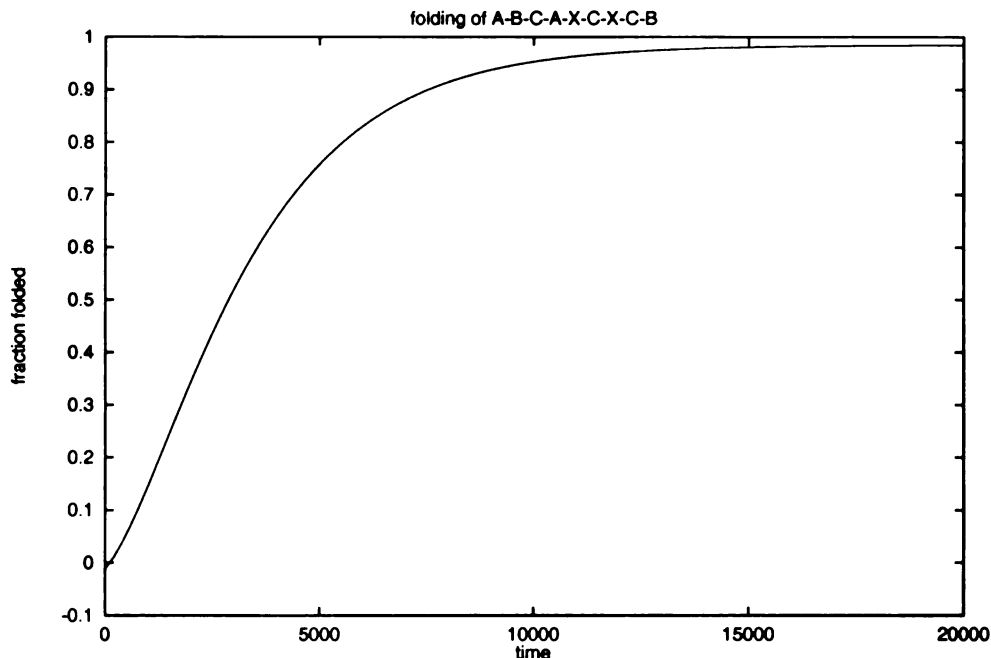
Figure 2.14: The fraction native as a function of time for ABCAXCXCB It is not necessary to fit this curve to a function, because we know that it is a sum of exponentials. The rate constants of the exponentials are given by the inverses of the eigenvalues. The initial value of each exponential is determined by the component of its eigenvector in the direction of the native conformation, multiplied by the initial amplitude of that eigenvector.

is very short relative to the slowest eigenvalue, the folding is is well represented by a single rate constant. Most of the lag phase can be accounted for by only one other rate constant. Thus the onset of the native state can be described by a simple mass action kinetic equation.

When started from an ensemble with all configurations equally likely, the slowest eigenvector typically accounts for 99.8 % of the time course for the onset of the lowest energy conformation, over a range of contact energies. We calculate this number by summing the contributions from all eigenvectors that give rise to an increase in the

population of the native state, and determining what fraction of this total is due to the each eigenvector. Single exponential folding kinetics, as defined here, was found to occur in all three maximally compact nine-mer native conformations, and for a variety of kinetic parameters ($H = 1, .5, .2, .1, \eta = .1, .2, .5, 1.$)

Bibliography

[1] Sophie E. Jackson. How do small single-domain proteins fold? *Folding and Design*, 3(4):R81–R91, August 1998.

[2] Randall E. Burton, Jeffrey K. Myers, and Terrence G. Oas. Protein folding dynamics: quantitative comparison between theory and experiment. *Biochemistry*, 37(6):5337–5343, 1998.

[3] P. S Kim and R. L. Baldwin. Intermediates in the folding reactions of small proteins. *Annual Review of Biochemistry*, 59:631–60, 1990.

[4] Alan R. Fersht. *Structure and Mechanism in Protein Science*. Freeman, 1999.

[5] Walter Englander. Protein folding intermediates and pathways studied by hydrogen exchange. *Annual Review of Biophysics and Biomelcular Structure*, 29:213–238, 2000.

[6] C. Levinthal. ? *J. Chim. Phys*, 85:44, 1968.

[7] Hans Eyring. ? *Chemical Reviews*, 17:65, 1935.

[8] Andreas Matouschek, James T. Kellis Jr, Luis Serrano, Mark Bycroft, and Alan R. Fersht. Transient folding intermediates characterized by protein engineering. *Nature*, 346:440–445, August 1990.

[9] V.S. Pande, A.Yu. Grosberg, T. Tanaka, and D.S. Rokhsar. Pathways for protein folding: is a new view needed? *Current Opinion in Structural Biology*, 8(1):68–79, 1998.

[10] R. E. Burton, G. S. Huang, M. A. Daugherty, T. L. Calderone, and T. G. Oas. The energy landscape of a fast-folding protein mapped by ala-¿gly substitutions. *Nature Structural Biology*, 4(4):305–310, 1997.

[11] A. R. Fersh. Transition-state structure as a unifying basis in protein-folding mechanisms: contact order, chain topology, stability, and the extended nucleus mechanism. *Proceedings of the National Academy of Sciences*, 87(4):1525–1529, 2000.

[12] Joseph D. Bryngelson, José Nelson Onuchic, Nichoas D. Socci, and Peter G. Wolynes. Funnels, pathways and the energy landscape of protein folding: a synthesis. *Proteins:*, 21:167–195, 1995.

[13] Ken A. Dill. Theory for the folding and stability of globular proteins. *Biochemistry*, 24(6):1501–1509, 1985.

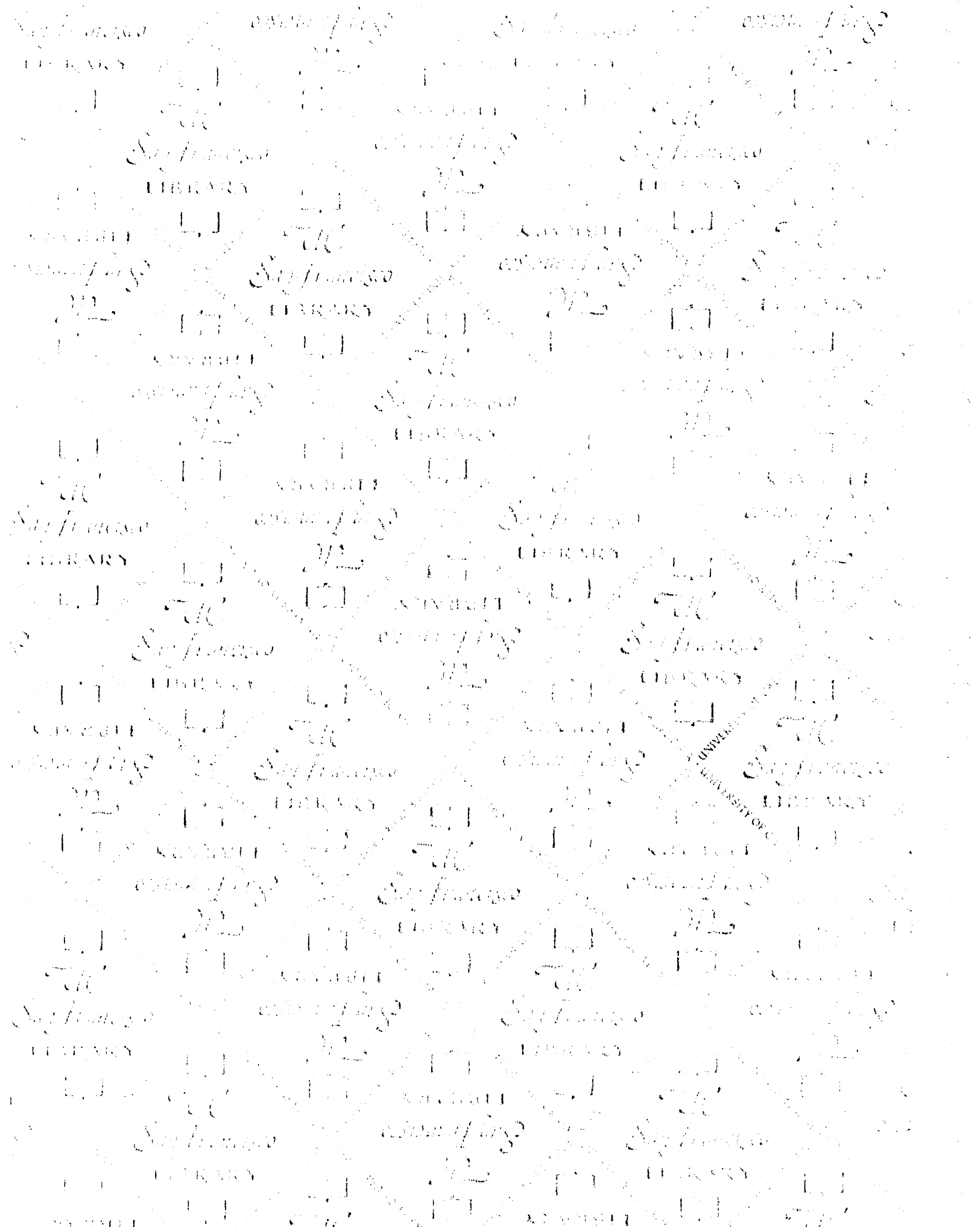[14] Joseph D. Bryngelson and Peter G. Wolynes. Spin glasses and the statistical

mechanics of protein folding. *Proceedings of the National Academy of Sciences*, 84(21):7524–7528, 1987.

[15] Ken A. Dill and Hue Sun Chan. From Levinthal to pathways to funnels. *Nature Structural Biology*, 4(1):10–19, January 1997.

[16] Ken A. Dill. Polymer principles and protein folding. *Protein Science*, 8(6):1166–80, June 1999.

[17] H. Taketomi, Ueda Y., and N. Go. Studies on protien folding, unfolding and fluctuations by computer simulations. *International Journal of Peptide and Protein Research*, 7:445–459, 1975.

[18] E. I. Shakhnovich, V. I. Abkevich, and O. B. Ptitsyn. Conserved residues and the mechanism of protein folding. *Nature*, 379:96–98, 1996.

[19] R. Miller, C.A. Danko, M.J. Fasolka, A.C. Balazs, H.S. Chan, and K.A. Dill. Folding kinetics of proteins and copolymers. *Journal of Chemical Physics*, 96(1):768–80, 1992.

[20] D. K. Klimov and D. Thirumalai. Lattice models for proteins reveal multiple folding nuclei for nucleation-condensation mechanism. *Journal of Molecular Biology*, 282:471–492, 1998.

[21] A. Kolinsk and J Skolnick. Monte carlo simulations of protein folding. i. lattice model and interaction scheme. *Proteins*, 18(4):338–352, 1994.

[22] KA Dill, S Bromberg, K Yue, KM Fiebig, DP Yee, PD Thomas, and HS Chan. Principles of protein folding—a perspective from simple exact models. *Protein Science*, 4(4):561–602, 1995.

[23] Shoji Takada. *Gō*-ing for the prediction of protein folding mechanisms. *Proceedings of the National Academy of Sciences*, 96(21):11698–11700, 1999.

[24] Victor Muñoz, Eric R. Henry, James Hofrichter, and William A. Eaton. A simple model for calculating the kinetics of protein folding from three-dimensional structures. *Proceedings of the National Academy of Sciences*, 96:11311–11316, September 1999.

[25] E. Alm and D. Baker. Prediction of protein-folding from free-energy landscapes derived from native structures. *Proceedings of the National Academy of Sciences*, 96:11305–11310, September 1999.

[26] Oxana V. Galzitskaya and Alexie V. Finkelstein. A theoretical search for folding/unfolding nuclei in three-dimensinal protein structures. *Proceedings of the National Academy of Sciences*, 96(20):11299–11304, September 1999.

[27] H. S. Chan. Modelling protein folding by monte carlo dynamics: chevron plots, chevron rollover, and non-arrhenius kinetics. In P. Grassberger, G. T. Barkema, and W. Nadler, editors, *Monte Carlo Approach to Biopolymers and Protein Folding*, pages 29–44. World Scientific, 1998.

[28] M. Cieplak, and J. Karbowski M. Henkel, and J. R. Banavar. Master equation approach to protein folding and kinetic traps". *Physical Review Letters*, 80:3654–3657, 1998.

[29] Richard C. Tolman. *The Principles of Statistical Mechanics*. Dover, New York, 1979.

[30] Martin J. Parker and Susan Marqusee. The cooperativity of burst phase reactions explored. *Journal of Molecular Biology*, 293:1195–1210, 1999.

[31] Sepideh Khorasanizadeh, Iain D. Petrs, and Heinrich Roder. Evidence for a three-state model of protein sfolding form kientics analysis of ubiquitin variants with altered core residues. *Nature Structural Biology*, 3(2):193–205, 1996.

[32] T. R. Sosnick, L. Mayne, and S. W. Englander. Molecular collapse: the rate-limiting step in two-state cytochrome c folding. *Proteins*, 24(4):413–426, 1996.

[33] Sophie E. Jackson and Alan R. Fersht. Folding of chymotrypsin inhibitor 2. 1. evidence for a two-state transition. *Biochemistry*, 30:10428–10435, 1991.

[34] Mikael Oliveberg. Alternative explanations for "multistate" kinetics in proteins folding: transient aggregation and changing transition-state ensembles. *Accounts of Chemical Research*, 31:765–772, 1997.

[35] D. E. Otzen, O. Kristensen, M. Procotor, and M Oliveberg. Structural changes

in the transition state of protein folding: alternative interpretations of curved chevron plots. *Biochemsitry*, 38:6499–6511, 1999.

[36] C. R. Matthews. Pathways of protein folding. *Annual Review of Biochemistry*, 62:653–683, 1993.

[37] Yaoqi Zhou and Martin Karplus. Interpreting the folding kinetics of helical proteins. *Nature*, 401:400–402, September 1999.

[38] Y Duan and P. A. Kollman. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science*, 282(5389):740–744, 1998.

[39] N.V. Dokholyan, S.V. Buldyrev, H.E. Stanley, and E.I. Shakhnovich. Identifying the protein folding nucleus using molecular dynamics. *Journal of Molecular Biology*, 296(5):1183–1188, 2000.

[40] H. S. Chan and K. A. Dill. Transition states and folding dynamics of proteins and heteropolymers. *Journal of Chemical Physics*, 100:9238–9257, 1994.

[41] H.J. Hilhorst and J.M. Deutch. Analysis of monte carlo results on the kinetics of lattice polymer chains with excluded volume. *Journal of Chemical Physics*, 63(12), 1975.

[42] H. Boots and J.M. Deutch. Analysis of the model dependence of monte carlo

results for the relaxation of the end-to-end distance of polymer chains. *Journal of Chemical Physics*, 67(10), 1977.

[43] N. G. van Kampen. *Stochastic Processes in Physics and Chemistry*. Elsevier, Amsterdam, 1997.

[44] F. E. Cohen, M. J. Sternberg, D. C. Phillips, I. D. Kuntz, and P. A. Kollman. A diffusion-collision-adehsion model for the kinetics of myoglobin refolding. *Nature*, 286(5773):632–634, 1980.

[45] Victor Muñoz, Eric R. Henry, James Hofrichter, and William A. Eaton. A statistical mechanical model for β-hairpin kinetics. *Proceedings of the National Academy of Sciences*, 95(11):5872–5879, May 1998.

[46] Robert Zwanzig. Two-state models of protein folding kinetics. *Proceedings of the National Academy of Sciences*, 94:148–150, 1997.

[47] William H. Press, Saul A. Teukolsky, Willima T. Vetterling, and Brian P. Flannery. *Numerical Recipes in C*, chapter 11. Cambridge University Press, 1992.