Diffusion-Collision Model for the Folding Kinetics of Myoglobin

D. Bashford,[1] F.E. Cohen,[2] M. Karplus,[3] I.D. Kuntz,[2] and D.L. Weaver[1]
[1]*Department of Physics, Tufts University, Medford, Massachusetts 02155;* [2]*Department of Pharmaceutical Chemistry, University of California, San Francisco, San Francisco, California 94143;* [3]*Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138*

ABSTRACT **The diffusion-collision model has been used to analyze the folding kinetics of myoglobin. The microdomains, which are the basic units that coalesce during the folding, are identified with the helices and the stabilizing contacts between helices are determined from the native structure. Both association and dissociation reactions are included and a range of stabilization parameters is investigated to determine the variation in overall rate and the relative contributions made by different intermediates during the folding process. In a comparison of folding to the native state and to the midpoint of the folding transition (i.e., 50% native protein at the completion of the reaction) significant differences in the contributing intermediates are found.**

Key words: protein-folding, multiple pathways

INTRODUCTION

Most approaches to the dynamics of protein folding have focused on mechanisms that avoid a random search through all possible structures and explain the experimentally determined folding times of milliseconds to minutes.[1] Since detailed measurements delineating the mechanism are lacking, theoretical analyses are needed. They can indicate the range of possibilities and suggest what to look for in experimental studies. One proposed folding mechanism, the diffusion-collision model,[2] postulates that an elementary step in protein folding is the collision between two possibly unstable quasiparticles called microdomains. These microdomains, which may be portions of incipient secondary structure (α-helices or β-strands) or hydrophobic clusters, reduce the search problem by dividing the protein into a number of segments (small relative to the total number of amino acid residues), each of which can fold and unfold rapidly. The overall dynamics of folding is then governed by a set of diffusion equations that describe the relative motion of the microdomains and their aggregates, and by boundary conditions which couple the diffusion equations through the collision and coalescence of the microdomains. A model of this type cannot address the finer details of folding that involve the atomic motions, but focuses instead on the larger scale processes that determine some aspects of folding times and the presence of experimentally measurable intermediates.

Folding is assumed to involve the diffusive motion of the microdomains, which are subject to random, external forces due to frequent collisions with solvent molecules, and well-defined internal forces (covalent bonds, hydrophobic interactions, H-bonds, etc.) that provide stabilizing interactions. Under renaturing conditions, collisions between microdomains sometimes lead to their coalescence into microdomain pairs and so on into larger microdomain aggregates that eventually result in the formation of the native structure. The diffusion-collision model reduces the dynamics of the folding process from consideration of individual amino acids in the polypeptide chain to consideration of the properties of microdomains. A Brownian dynamics study[3] has shown that the rate of the elementary two-body step of the diffusion-collision process can be estimated by the analytically soluble classical diffusion model used in the earlier analyses.[4]

Another mechanism proposed for folding is the random-search nucleation and chain propagation model.[5] Such a mechanism has been argued against by Baldwin[1] on both experimental and theoretical grounds. It has recently been elaborated and modified by Gō and Abe[6] into the noninteracting local structure or "growth-merge" model. An "embryo," unstable by itself and possibly equivalent to a microdomain, serves as a nucleus for the folding of polypeptide chain segments. This embryo grows until it merges with another growing nucleus. The growth-merge mechanism clearly has points in common with the diffusion-collision model; a difference appears to be in whether the growth of a single embryo (microdomain) or the coming together (diffusion and coalescence) of two or more of these elements governs the overall kinetics of the folding process. In both models, nearest neighbor microdomains are most likely to coalesce first (everything being equal) whether by growth or

Received February 23, 1988; revision accepted August 3, 1988.

Dr. Donald Bashford's current address is Department of Chemistry, Harvard University, Cambridge, MA 02138. Address reprint requests there.

via collisions. This is in accord with the distribution of contacts observed between secondary structural elements in proteins.[7,8]

For myoglobin, which is the primary focus of this paper, the clear division of the protein into α-helices and connecting random coil or loop segments makes a growth-merge mechanism appear unlikely so that the diffusion-collision model is the natural one to apply. In their pioneering study of myoglobin, Ptitsyn and Rashin,[9] who were not concerned with folding kinetics per se, clearly suggested that the helices were the folding units. In what follows, we apply the diffusion-collision model to myoglobin, extending and correcting an earlier study of this system.[10] Methods describes the procedure for evaluation of the parameters involved in the elementary steps and the formulation and solution of the first-order differential equations that govern the folding process. For the application to myoglobin, the microdomains and the kinetic parameters appropriate to them have to be determined; this is done in the section entitled Myoglobin Model and Parameters. The results of the myoglobin calculations are presented and discussed in Results and Discussion.

METHODS

In this section, we describe first how to relate the parameters for the elementary association and dissociation reactions to the properties of the microdomains. We then introduce the chemical kinetic approximation and show how it can be used to combine the elementary steps into a mechanism for the folding of the entire protein.

Association Reaction

The motions of the microdomains that lead to coalescence are assumed to be dominated by solvent drag; systematic forces due to the internal potential are taken to be negligible. Under these conditions, the forward transitions are governed by a set of diffusion equations which describe the motion of microdomains and microdomain aggregates (to be identified with the seven interacting helices in myoglobin, see Myoglobin Model and Parameters), and by boundary conditions which introduce the collision and coalescence of the aggregates. Analytical and numerical simulations[11] have shown that the kinetics of aggregate formation and dissociation may be approximated by individual quasi two-body rates and that the forward (folding) two-body rates may be estimated from a rate formula that depends only on measurable physical quantities associated with individual states. The movement of a pair of microdomains is approximated by a diffusion equation of the form

$$\frac{\partial \rho}{\partial t} = D\nabla^2 \rho \qquad (1)$$

where ρ is the relative position probability density of the two microdomains and D is the relative diffusion coefficient. As a simplification, we make the idealization that the microdomains are spheres connected by a perfectly flexible "string." Model calculations and simulations of helices connected by random coil chains suggest that this leads to the correct orders of magnitude, in part due to a cancellation of errors.[3] The maximal radial separation, R_{max}, of the microdomains is limited by the length of the string (the polypeptide chain between microdomains); the minimum separation, R_{min}, is assumed to be the sum of the radii of the two spheres.

The probability, p(t), that the microdomain pair is in an unbonded state at a time, t, is given by the integral of ρ between the minimum and maximum limits, that is,

$$p(t) = \int_{R_{min}}^{R_{max}} \rho(\vec{x},t)\, d^3x \qquad (2)$$

Previous work[4] has shown that the probability p(t) is well approximated by a simple relaxation time expression

$$p(t) \cong e^{-t/\tau_b} \qquad (3)$$

where τ_b is the mean coalescence time, which has the form[12]

$$\tau_b = \frac{G}{D} + \frac{LV(1-\beta)}{\beta DA} \qquad (4)$$

The geometrical parameters in Equation 4 are V, the volume of diffusion space available for the relative motion, taken as the annular volume between concentric spheres of radii R_{min} and R_{max}; A, the relative surface area available for collisions, taken as the area of a sphere of radius R_{min}; and G which is equal to (defining $\epsilon = R_{min}/R_{max}$)

$$G = \frac{R_{max}^2}{3}\,\frac{1 - (9/5)\epsilon + \epsilon^3 - (1/5)\epsilon^6}{\epsilon(1-\epsilon^3)} \qquad (5)$$

The quantity L in Equation 4 is a parameter with units of length that can be derived by using a stochastically switching reaction boundary condition with Equation 1.[12] The resulting form for L is

$$\frac{1}{L} = \frac{1}{R_{min}} + \alpha\,\frac{\alpha R_{max}\tanh[\alpha(R_{max}-R_{min})] - 1}{\alpha R_{max} - \tanh[\alpha(R_{max}-R_{min})]} \qquad (6)$$

where

$$\alpha \equiv \left(\frac{\lambda_1+\lambda_2}{D}\right)^{1/2}$$

The quantity λ_1 is the rate for the switching of a nonassociated microdomain pair from a state in which both microdomains are correctly folded and oriented for association to any other state, and λ_2 is the rate for the reverse reaction. In terms of λ_1 and λ_2, one has $\beta \equiv \lambda_2/(\lambda_1+\lambda_2)$, the probability that the two microdomains are in the folded, oriented state when they collide so that there is no barrier to association; alternatively, β can be defined to include activation energy and orientational effects.[4]

A useful parameter for discussion of the folding kinetics is the overall rate of the helix-coil transition, which is the essential element in determining whether a chain segment is in a folded state that can coalesce. This rate is given by $\lambda_1+\lambda_2 \equiv 1/\tau_c$. With this definition of τ_c we have $\alpha = 1/\sqrt{D\tau_c}$. Experimental[13-17] and theoretical[18] studies suggest that helix coil transitions occur in times of the order of 10^{-7} to 10^{-10} seconds, so we have taken $\tau_c = 10^{-8}$ seconds for the present calculations.

The relative diffusion constant D is given by

$$D = [k_BT/6\pi\eta](1/R_1 + 1/R_2) \qquad (7)$$

where η is the viscosity of water at absolute temperature T, and k_B is the Boltzmann constant.

The radii of the interacting microdomains, R_1 and R_2, are estimated by a simple approximation. We sum the van der Waals volumes of the atoms in a microdomain (i.e., taking atomic sphere volumes obtained from the van der Waals radii given in the CHARMM program[19] and neglecting overlap), multiply by $3\sqrt{2}/\pi$, the packing ratio for spheres in tetrahedral packing (as an approximation to the actual irregular packing) and calculate the radius of a sphere with the same volume. The sum of the radii, R_1 and R_2, is used for R_{min} in Equations 3–6. The individual microdomain radii calculated by this method are given in Table III.

R_{max} between two microdomains is the fully extended length of the connecting chain plus the distances on each end between the centroid of the microdomain's atoms and the attachment point of the chain (taken as the last α-carbon before the chain). The center and connection points are taken from the crystal structure. The length of the extended connecting chain is taken as 3.5 Å per residue for random coil segments. If there are microdomains along the connecting chain, their contribution to the length is the distance between the random coil connection points on each end. The R_{max} values calculated by this method are given in Table IV.

Since there is little experimental evidence to indicate the correct value for β, it is taken as a free parameter of the theory. We have made calculations for a series of β values using the convention that β for a pair of microdomains, β_{ij}, is equal to

$$\beta_{ij} = \beta_i\beta_j \qquad (8)$$

where β_i is the probability of finding microdomain i in the helical and correctly oriented state (see above).

Dissociation Reaction

For the rate at which microdomain pairs or larger microdomain clusters dissociate, we use the formula

$$\frac{1}{\tau_{ij}} = \upsilon e^{-fA_{ij}/k_BT} \qquad (9)$$

where A_{ij} is the change in contact surface area as calculated by the Lee and Richards algorithm[20,21] when microdomains i and j coalesce; and f = 80 cal/mole/Å^2, is an estimate of the free energy change due to the burial of hydrophobic surfaces.[21,22] The value of 80 is appropriate (see Table I of ref. 21) for the "contact" area used here and corresponds to the value of 23 for "accessible" area. For the two-microdomain case, agreement between the ratio of forward and backward rate constants and the classical expression for the equilibrium constant would require putting $\upsilon = \tau_b^{-1} Q_d/Q_a$, where τ_b^{-1} is defined in Equation 3, and Q_d/Q_a is the ratio of the configuration space volumes of the dissociated and associated states (i.e., the appropriate partition functions). Since an evaluation of multidomain configurational volumes is beyond the level of detail of the present study, we have taken υ as a single parameter to be determined by the equilibrium properties of the system as a whole. Under native solution conditions the value of υ is adjusted arbitrarily to have 95% of the total protein concentration in the correctly folded form at equilibrium; any value between 95% and 100% would yield very similar results. To investigate the kinetics in the transition region,[1] the fully folded protein was required to be present at 50% concentration. Equation 9 is clearly somewhat arbitrary; e.g., the proportionality of area and energy is approximately valid for van der Waals interactions, but not necessarily for hydrophobic or charge interactions.

Chemical Kinetics Approximation

Because the solution of the diffusion equation leads to a simple exponential dependence of probability on time (Eq. 3), we can formulate the folding rate as a chemical kinetics problem.[11] This results in a simple procedure for treating a system composed of many states. Equation 3 corresponds to the solution of the unimolecular rate equation

$$\frac{dp}{dt} = Rp \qquad (10)$$

where $R = -(1/\tau_b)$. The generalization to the many state problem is

$$\frac{dp_i}{dt} = \sum_{j=1}^{N} R_{ij}p_j \qquad (i = 1, 2 \ldots N) \qquad (11)$$

where $p_i = p_i(t)$ is the probability of finding the molecule in the i_{th} state at time t, and R_{ij} is the transition probability per unit time from the j^{th} to the i^{th} state;

TABLE I. Helix-Helix Bonds in Apomyoglobin*

Helices	Residues	Solvent contact area loss (Å)
A	4–17	185
H	125–148	
B	21–35	122
D	52–57	
B	21–35	230
E	59–79	
B	21–35	203
G	101–117	
F	83–96	156
H	125–148	
G	101–117	253
H	125–148	

*Apomyoglobin helices and helix-helix area losses used in Equation 9 for the backward rates.

the states for myoglobin are defined below (Myoglobin Model and Paramters). The set of the equations given by Equation 11 can be solved by finding the eigenvalues and eigenvectors of an (N−1) by (N−1) matrix that is closely related to the R matrix in Equation 11. This is much simpler than solving the set of coupled partial differential equations that would result from the direct application of the diffusion equation to the many state problem. Details of the solution are given in the Appendix.

In an attempt to apply the diffusion-collision model to myoglobin, Cohen et al.[10] used second-order rate equations of the form

$$\frac{dA}{dt} = K(A)(B) \tag{12}$$

where (A) is the "concentration of microdomain A" and similarly for B. If the microdomains A and B were not linked together by a chain and each A was free to interact with any other B in solution, this equation would be correct and the frequency of collisions would be proportional to the product of (A) and (B). However, this is not the case for the folding of a single protein chain where a given A can collide only with a B in the same chain to give a folded structure. Furthermore, it is not really valid to speak of the "concentration of microdomain A" as if it were independent of B. This second point may seem trivial in the two microdomain case where A and B disappear at the same rate. It becomes important when systems of several microdomains are considered, as in the overall folding of myoglobin.

MYOGLOBIN MODEL AND PARAMETERS

To apply the diffusion-collision model to the folding of myoglobin, one must identify the microdomains and the interactions that are involved in producing the major features of the folding dynamics. In its native conformation, myoglobin (Mb) consists of eight α-helical regions (labeled by letters A to H starting from the N-terminal end) connected by nonhelical chains of various lengths. Helices and groups of helices are assumed to constitute the microdomains. The heme is not included in the calculation, which thus is a model for the folding of apomyoglobin rather than myoglobin. Helix–helix interactions are required for the stability of both the tertiary structure and the helices themselves. The details of these interactions, which we will call "bonds," are not specified but are assumed to be dominated by van der Waals and hydrophobic interactions, though dipole and other charge interactions may also be important.

Although quantitative experimental data are not available, it is clear that the helix-helix contacts are primarily responsible for the stability of the individual helices and the tertiary structure of the molecule; i.e., in the absence of such contacts the helices are in equilibrium with their "random coil" forms. Measurements on apomyoglobin indicate that the heme also is involved in stabilizing the native structure,[23] but this contribution is neglected in the present calculations.

We use the loss of contact surface area upon association as the measure of helix-helix interaction. Richmond and Richards[21] have calculated the contact surfaces of the helix-helix interactions, and have identified six contacts among seven of the eight helices (the C helix appears to act as a connector only) as being significant in terms of loss of accessible area to solvent. The contacts are GH, BE, BG, AH, FH, and BD in decreasing order of helix-helix contact area. These are summarized in Table I. The results of our area calculation are in accord with those of Richmond and Richards.

The contact surface area analysis identifies a reasonable list of helix-helix pairs and their relative interaction strengths. The results are consistent with earlier studies of globin sequences[24,25] and the more recent work of Lesk and Chothia.[26] They studied the structural and sequence homologies of nine globins and found that the tertiary structure is determined principally by helix-helix and helix-heme packing and identified 59 residue positions involved in these contacts. Thirty-one of these they classify as "buried" because they have accessible surface areas of 10 Å^2 or less. Although the identity of the residues found in these 31 buried contact positions varies, they are nearly always hydrophobic in character. All of the 31 buried residues of Lesk and Chothia except those involved with the heme (since we consider apomyoglobin) are found to have area loss on association, and 50% of the total area loss comes from these residues. This suggests that the present model may be of general significance for the globins. A specific difference is that the helix B–helix D interaction which is found

TABLE II. Myoglobin States*

State	Microdomains	State	Microdomains
1	A B C D E F G H	33	A B C D E F GH
2	B C D E F G AH	34	B C D E F AGH
3	A C E F G H BD	35	A C E F BD GH
4	C E F G AH BD	36	C E F BD AGH
5	A C D F G H BE	37	A C D F BE GH
6	C D F G AH BE	38	C D F BE AGH
7	A C F G H BDE	39	A C F BDE GH
8	C F G AH BDE	40	C F BDE AGH
9	A C D E F H BG	41	A C D E F BGH
10	C D E F AH BG	42	C D E F ABGH
11	A C E F H BDG	43	A C E F BDGH
12	C E F AH BDG	44	C E F ABDGH
13	A C D F H BEG	45	A C D F BEGH
14	C D F AH BEG	46	C D F ABEGH
15	A C F H BDEG	47	A C F BDEGH
16	C F AH BDEG	48	C F ABDEGH
17	A B C D E G FH	49	A B C D E FGH
18	B C D E G AFH	50	B C D E AFGH
19	A C E G BD FH	51	A C E BD FGH
20	C E G BD AFH	52	C E BD AFGH
21	A C D G BE FH	53	A C D BE FGH
22	C D G BE AFH	54	C D BE AFGH
23	A C G BDE FH	55	A C BDE FGH
24	C G BDE AFH	56	C BDE AFGH
25	A C D E BG FH	57	A C D E BFGH
26	C D E BG AFH	58	C D E ABFGH
27	A C E BDG FH	59	A C E BDFGH
28	C E BDG AFH	60	C E ABDFGH
29	A C D BEG FH	61	A C D BEFGH
30	C D BEG AFH	62	C D ABEFGH
31	A C BDEG FH	63	A C BDEFGH
32	C BDEG AFH	64	C ABDEFGH

*The folding states of apomyoglobin with numbering as described earlier (see Myoglobin Models and Parameters). A notation such as "B" indicates that the B helix is free; "BDE" indicates that helices B, D, and E are bonded together.

in myoglobin is not among the interactions generally found among the globins. Also, there are a number of other helix-helix contacts in myoglobin, AE, BC, EF, AG, CG, AB, and EH. However, they bury less surface area than the ones we have included. For example, the interaction of helix E with helix B, and helix A with helix H are stronger than those of helix A and helix E. These weaker interactions have been neglected in order to obtain a relatively simple model.

Given the above structural results, the observed bonding pattern of folded myoglobin is assumed to be the result of two-body-like interactions among the seven interacting helices with the kinetics governed by the diffusion-collision model. In the unfolded state, helices can appear transiently but are unstable. These fluctuating regions of the polypeptide chain are called one-helix microdomains. They diffuse together and "bond" to form more stable multihelix microdomains. This bonding is reversible but under renaturing solution conditions, the system progresses by forming more stable multihelix microdomains, until the native structure is reached. If the overall "state" of the molecule is defined in terms of the presence or absence of its possible "bonds," folding can be described as a set of transitions towards states with increased numbers of bonds.

To label the possible states of the myoglobin folding model, the two-body interactions between helices have been assigned a binary representation (0 = no "bond", 1 = "bond" formed). The folding state is then represented by a six-digit binary number with the digits corresponding to the "bonds" between, respectively, helices A–H, B–D, B–E, B–G, F–H, and G–H. Thus, 000001 represents the state with the A–H interaction and no others, and 100000 represents the state with the G–H interaction and no others. The state is labeled by adding one to the decimal equivalent of the binary number so formed. For example, 000000 is state 1, 111111 is state 64, and 110101 is state 54. Table II shows the labeling for each state.

The geometric and transport parameters corresponding to each of the helices and helix clusters are listed in Table III. Table IV is a complete listing of the forward transitions including the R_{max} for each microdomain pair used in Equation (5).

RESULTS AND DISCUSSION

The set of rate equations describing myoglobin folding has been solved for two different solution conditions. In one case the final folded state, state 64, represented 95% of the system at equilibrium; this is referred to as $P_{eq} = 0.95$. Other runs were made with state 64 equal to 50% at equilibrium, the midpoint of the folding-unfolding transition. These different limiting results were obtained by varying the parameter υ which scales the dissociation rate (see Eq. 9). In addition, the absolute and relative values of β were adjusted to correspond to different coalescence properties and to different choices for the variation of the coalescence properties with the size of the microdomain cluster. The simulations are summarized in Table V and the results are displayed graphically in Figures 1–6. Only intermediate states with maximum populations of 0.02 or more are shown in the figures, although all 64 states were included in the calculations.

We discuss the first case in detail to indicate the nature and significance of the results ($P_{eq} = 0.95$ with $\beta = 1$ for every helix and cluster of helices). Since all β values are equal to unity, the geometry of the helices and their contacts are most important in determining the folding kinetics. Figure 1 shows the results. The only significantly (>0.02) populated one-bond state is 33(GH), which reaches a maximum population of 0.06 after 4 ns and falls below 0.01 by 100 ns. On a nearest-neighbor basis, one would have expected BD and GH to form most easily, followed by BE and FH. However, only GH reaches significant values and the others coalesce rapidly to form the

TABLE III. Myoglobin Microdomains*

Label	Radius (Å)	Mass (amu)	D ($Å^2$/ns)
A	9.72	1,637.00	33.757507
B	9.58	1,545.00	34.251205
C	7.17	696.00	45.784035
D	6.97	658.00	47.085987
E	9.98	1,810.00	32.897316
F	9.27	1,463.00	35.417522
G	10.28	1,887.00	31.919765
H	11.22	2,600.00	29.249708
AH	13.26	4,237.00	24.750387
BD	10.68	2,203.00	30.728125
BE	12.33	3,355.00	26.626244
BDE	13.03	4,013.00	25.191162
BG	12.53	3,432.00	26.194887
BDG	13.21	4,090.00	24.843840
BEG	14.36	5,242.00	22.858690
BDEG	14.89	5,900.00	22.048012
FH	13.02	4,063.00	25.202528
AFH	14.62	5,700.00	22.442822
GH	13.57	4,487.00	24.182869
AGH	15.06	6,124.00	21.786383
BGH	15.01	6,032.00	21.870228
ABGH	16.26	7,669.00	20.185204
BDGH	15.49	6,690.00	21.185028
ABDGH	16.68	8,327.00	19.681376
BEGH	16.35	7,842.00	20.070704
ABEGH	17.43	9,479.00	18.834194
BDEGH	16.76	8,500.00	19.577797
ABDEGH	17.79	10,137.00	18.448738
FGH	14.88	5,950.00	22.054678
AFGH	16.15	7,587.00	20.318567
BFGH	16.10	7,495.00	20.381901
ABFGH	17.21	9,132.00	19.074205
BDFGH	16.53	8,153.00	19.858906
ABDFGH	17.58	9,790.00	18.669357
BEFGH	17.29	9,305.00	18.982744
ABEFGH	18.26	10,942.00	17.974909
BDEFGH	17.66	9,963.00	18.585369
ABDEFGH	18.59	11,600.00	17.653437

*The properties of the microdomains. Sums of the radii are used for R_{min} in Equations 5 and 6 for the forward rates.

populated two-bond states 7(BDE) and 49(FGH). State BDE rises rapidly to 0.15 after 14 ns and gradually falls below 0.01 by 90 ns. State FGH is the most populated intermediate; it rises to a peak of 0.30 after 20 ns and falls below 0.01 by 150 ns. These two clusters centered on helix B and on helix H are expected on the basis of nearest-neighbor interactions since the diffusion times are smaller and bonding is assumed to occur on each collision. Both clusters have rather long lifetimes.

The only significantly populated intermediate with four bonds is state 55 (BDE-FGH), which is composed of the two principal two-bond clusters, which do not interact with one another. This state reaches a maximum population of 0.14 at 28 ns, and falls below 0.01 at 130 ns, again a rather long lifetime. There is one significantly populated five-bond cluster, state 62 (ABEFGH), which lacks the BD-bond and slowly rises to 0.048 at equilibrium; i.e., the equilibrium folded state consists of 62 (0.05) and 64 (0.95). The BD bond, which is missing in 62, has the smallest solvent contact area loss of all the bonds. The folded state 64 (ABDEFGH) with all helices paired begins to rise after a lag of a few nanoseconds while the initial pairing takes place. It reaches its equilibrium value of 0.95 in an exponential manner with a folding time of 48 ns.

The second calculation differs from the first in that the folding reactions among the helices and helix clusters is slowed down (β = 0.01 for all microdomains) while maintaining P_{eq} = 0.95. The one bond state, 33 (GH), observed in Figure 1 becomes more

TABLE IV. Myoglobin Transitions*

Transition of states	Bond formed	Microdomains	R_{max} (Å)
1→2	AH	A + H	270.79
1→3	BD	B + D	65.05
2→4	BD	B + D	65.05
3→4	AH	A + H	205.99
1→5	BE	B + E	89.57
2→6	BE	B + E	89.57
5→6	AH	A + H	171.30
3→7	BE	BD + E	29.25
5→7	BD	BE + D	23.72
4→8	BE	BD + E	29.25
6→8	BD	BE + D	23.72
7→8	AH	A + H	171.30
1→9	BG	B + G	176.70
2→10	BG	B + G	83.69
9→10	AH	A + H	84.63
3→11	BG	BD + G	116.37
9→11	BD	BG + D	63.71
4→12	BG	BD + G	84.99
10→12	BD	BG + D	63.71
11→12	AH	A + H	84.63
5→13	BG	BE + G	89.26
9→13	BE	BGG + E	88.23
6→14	BG	BE + G	80.20
10→14	BE	BG + E	88.23
13→14	AH	A + H	84.63
7→15	BG	BDE + G	91.29
11→15	BE	BDG + E	34.80
13→15	BD	BEG + D	27.32
8→16	BG	BDE + G	81.16
12→16	BE	BDG + E	34.80
14→16	BD	BEG + D	27.32
15→16	QH	A + H	84.63
1→17	FH	F + H	98.42
2→18	FH	F→AH	95.05
17→18	AH	A + FH	172.57
3→19	FH	F + H	98.42
17→19	BD	B + D	65.05
4→20	FH	F→AH	95.05
18→20	BD	B + D	65.05
19→20	AH	A + FH	107.77
5→21	FH	F + H	98.42
17→21	BE	B + E	89.57
6→22	FH	F + AH	77.12
18→22	BE	B + E	86.83
21→22	AH	A + FH	73.08
7→23	FH	F + H	98.42
19→23	BE	BD + E	29.25
21→23	BD	BE + D	23.72
8→24	FH	F + AH	77.12
20→24	BE	BD→E	29.25
22→24	BD	BE + D	23.72
23→24	AH	A + FH	73.08
9→25	FH	F + H	188.98
17→25	BG	B + G	176.70
10→26	FH	F + AH	176.61
18→26	BG	B + G	83.61
25→26	AH	A + FH	82.13
11→27	FH	F + H	133.38
19→27	BG	BD + G	116.37
25→27	BD	BG + D	63.71
12→28	FH	F + AH	111.81

(continued)

TABLE IV. Myoglobin Transitions* (Continued)

Transition of states	Bond formed	Microdomains	R_{max} (Å)
20→28	BG	BD + G	84.91
26→28	BD	BG + D	63.71
27→28	AH	A + FH	82.13
13→29	FH	F + H	100.39
21→29	BG	BE + G	89.26
25→29	BE	BG + E	88.23
14→30	FH	F + AH	77.12
22→30	BG	BE + G	80.12
26→30	BE	BG + E	88.23
29→30	AH	A + FH	73.08
15→31	FH	F + H	100.39
23→31	BG	BDE + G	91.29
27→31	BE	BDG + E	34.80
29→31	BD	BEG + D	27.32
16→32	FH	F + AH	77.12
24→32	BG	BDE + G	81.08
28→32	BE	BDG + E	34.80
30→32	BD	BEG + D	27.32
31→32	AH	A + FH	73.08
1→33	GH	G + H	58.88
2→34	GH	G + AH	55.51
33→34	AH	A + GH	209.50
3→35	GH	G + H	58.88
33→35	BD	B + D	65.05
4→36	GH	G + AH	55.51
34→36	BD	B + D	65.05
35→36	AH	A + GH	144.70
5→37	GH	G + H	58.88
33→37	BE	B→E	89.57
6→38	GH	G + AH	55.51
34→38	BE	B + E	89.57
37→38	AH	A→GH	110.01
7→39	GH	G + H	58.88
35→39	BE	BD + E	29.25
37→39	BD	BE + D	23.72
8→40	GH	G + AH	55.51
36→40	BE	BD + E	29.25
38→40	BD	BE + D	23.72
39→40	AH	A + GH	110.01
9→41	GH	BG + H	57.45
33→41	BG	B + GH	174.25
10→42	GH	BG + AH	41.77
34→42	BG	B + AGH	38.90
41→42	AH	A + BGH	39.54
11→43	GH	BDG + H	57.58
35→43	BG	BD + GH	113.92
41→43	BD	BGH + D	67.90
12→44	GH	BDG + AH	41.45
36→44	BG	BD + AGH	40.20
42→44	BD	ABGH + D	70.37
43→44	AH	A + BDGH	38.54
13→45	GH	BEG + H	59.02
37→45	BG	BE + GH	86.81
41→45	BE	BGH + E	85.99
14→46	GH	BEG + AH	38.65
38→46	BG	BE + AGH	35.41
42→46	BE	ABGH + E	88.45
45→46	AH	A + BEGH	37.05
15→47	GH	BDEG + H	58.92
39→47	BG	BDE + GH	88.84
43→47	BE	BDGH + E	40.57

(continued)

TABLE IV. Myoglobin Transitions* (Continued)

Transition of states	Bond formed	Microdomains	R_{max} (Å)
45→47	BD	BEGH + D	31.75
16→48	GH	BDEG + AH	38.73
40→48	BG	BDE + AGH	36.37
44→48	BE	ABDGH + E	41.77
46→48	BD	ADEGH + D	32.76
47 + 48	AH	A + BDEGH	36.47
17→49	GH	G + FH	46.06
33→49	FH	F + GH	37.13
18→50	GH	G + AFH	49.85
34→50	FH	F + AGH	40.80
49→50	AH	A + FGH	176.37
19→51	GH	G + FH	46.06
35→51	FH	F + GH	37.13
49→51	BD	B + D	65.05
20→52	GH	G + AFH	49.85
36→52	FH	F + AGH	40.80
50→52	BD	B + D	65.05
51→52	AH	A + FGH	111.57
21→53	GH	G + FH	46.06
37→53	FH	F + GH	37.13
49→53	BE	B + E	89.57
22→54	GH	G + AFH	49.85
38→54	FH	F + AGH	40.80
50→54	BE	B + E	86.83
53→54	AH	A + FGH	76.88
23→55	GH	G + FH	46.06
39→55	FH	F + GH	37.13
51→55	BE	BD + E	29.25
53→55	BD	BE + D	23.72
24→56	GH	G + AFH	49.85
40→56	FH	F + AGH	40.80
52→56	BE	BD + E	29.25
54→56	BD	BE + D	23.72
55→56	AH	A + FGH	76.88
25→57	GH	BG + FH	48.45
41→57	FH	F + BGH	38.18
49→57	BG	B + FGH	141.12
26→58	GH	BG + AFH	42.91
42→58	FH	F + ABGH	40.64
50→58	BG	B + AFGH	39.86
57→58	AH	A + BFGH	39.74
27→59	GH	BDG + FH	49.82
43→59	FH	F + BDGH	38.89
51→59	BG	BD + FGH	80.79
57→59	BD	BFGH + D	69.68
28→60	GH	BDG + AFH	42.59
44→60	FH	F + ABDGH	40.78
52→60	BG	BD + AFGH	41.15
58→60	BD	ABFGH + D	71.16
59→60	AH	A + BDFGH	38.80
29→61	GH	BEG + FH	48.86
45→61	FH	F + BEGH	39.26
53→61	BG	BE + FGH	53.68
57→61	BE	BFGH + E	54.13
30→62	GH	BEG + AFH	39.78
46→62	FH	F + ABEGH	40.97
54→62	BG	BE + AFGH	36.36
58→62	BE	ABFGH + E	53.67
61→62	AH	A + BEFGH	37.76
31→63	GH	BDEG + FH	49.85
47→63	FH	F + BDEGH	39.84
55→63	BG	BDE + FGH	55.71
59→63	BE	BDFGH + E	41.38
61→63	BD	BEFGH + D	32.06
32→64	GH	BDEG + AFH	39.87
48→64	FH	F + ABDEGH	41.17
56→64	BG	BDE + AFGH	37.33
60→64	BE	ABDFGH + E	42.18
62→64	BD	ABEFGH + D	32.80
63→64	AH	A + BDEFGH	37.14

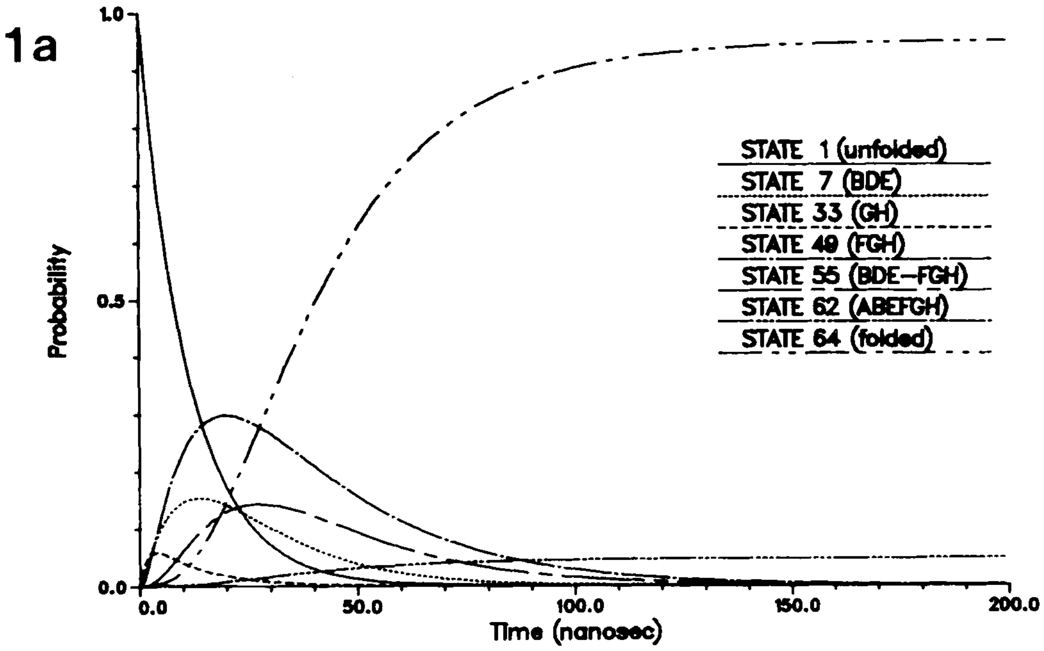
(continued)

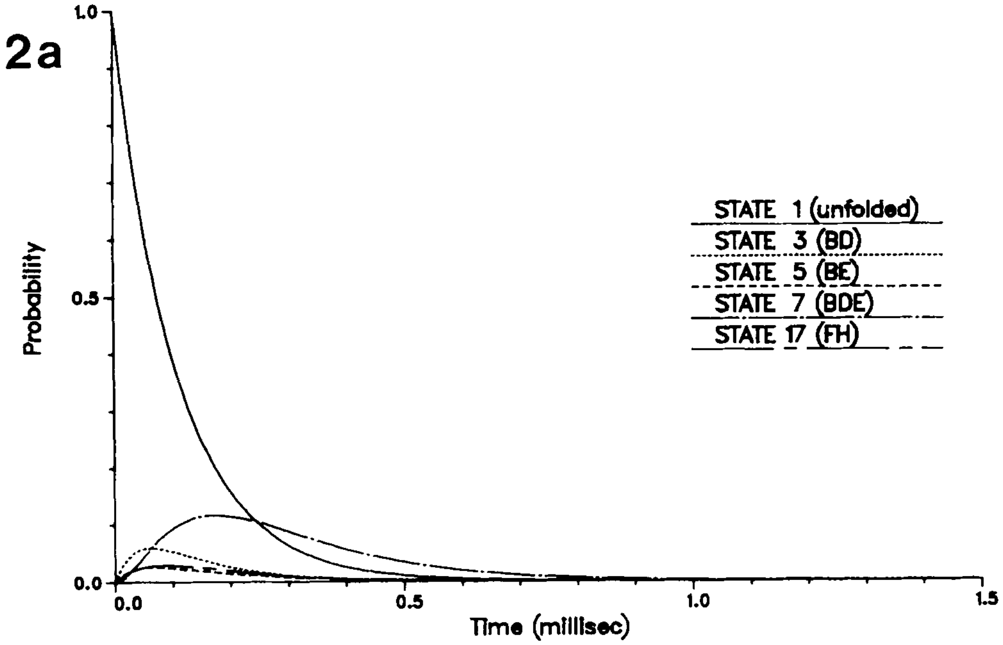
*The elementary folding steps. In each of these transitions one bond is formed. The R_{max} values are used in Equations 5 and 6 to calculate the forward rates.

prominent, rising to a larger value and having a longer lifetime (see Fig. 2). As a consequence, the most prominent two-bond and four-bond states—namely, 7 (BDE), 49 (FGH) and 55 (BDE-FGH)—rise to lower values, but are still quite prominent. Compared with $\beta = 1$ (Fig. 1) there are seven additional intermediate states which are significantly populated; they are 3 (BD), 5 (BE), 17 (FH), 39 (BDE-GH), 51 (BD-FGH), 56 (AFGH-BDE), and 63 (BDEFGH). In addition, the lifetimes of the intermediates are longer and the overall folding time is slowed down from a submicrosecond to a millisecond time scale.

It is expected that interacting helices are more stable than isolated ones; so the β values of more complex clusters of helices should be larger than those of isolated helices. Neglecting orientational effects, the individual β's for a cluster may be approximated by the coil-helix equilibrium constants of the helices in the clusters involved in the pairing. These equilibrium constants are governed by the helix propagation steps in a nucleation-propagation model of the helix-coil transition since internal contacts in the cluster insure the presence of a nucleus. Values of $\beta > 0.1$ are expected from experimentally determined propagation parameters.[27] An experimental prototype of such helix pairing is provided by glucagon aggregation in solution. Nuclear magnetic resonance (NMR) data[28] show that monomers are in the coil state and that trimer association takes place in the carboxy-terminal region, which is in a helical form, as in the crystal. However, the helical structure from Phe 6 to Tyr 13 present in the crystal trimers due to stabilizing contacts with other trimers is absent in solution. From ref. 27, the product of the helix propagation Parameters for the glucagon residues from Phe 6 to Tyr 13 is approximately 0.36.

A number of runs were carried out in which the value of β assigned to the microdomains increased as the microdomain clusters became more complex (see Table V). An extreme example is shown in Figure 3 in which $\beta = 0.01$ for individual helices and $\beta = 1$ for clusters of two or more helices. Most of the intermediates are now suppressed, and only the two helix clusters BDE and FGH appear, along with the state 62 (ABEFGH), which contributes 0.05 at equilibrium.





Figs. 1–6. The results of runs 1–6, respectively (see Table V). Probability (or relative concentration) of the folded, unfolded, and intermediate states is plotted as a function of time. All 64 states are present in the calculations, but only those rising above 0.02 are plotted. The notation used to indicate the states in the legends is explained earlier (see Myoglobin Model and Parameters and Table II). The symbol used is indicated under each state in the figure.

(*Figs. 2b through 6 appear on following pages*)

In this case the model resembles nucleation-controlled folding in which the first steps are much slower (the nucleation steps) than the subsequent (propagation) steps, so that the latter mostly are not observable. Figure 4 shows a more gradual increase of cluster stability, in which $\beta = 0.1$ for individual helices, $\beta = 0.5$ for two-helix clusters and $\beta = 1$ for clusters of more than two helices. The general behavior is similar to Figure 3, but there are additional intermediates that make larger contributions.

Many experimental folding studies are made near the midpoint of the transition,[29] where the folded state 64 (ABDEFGH) is present at a relative concentration of 0.5 at equilibrium. Figure 5 shows a run with $\beta = 1$ for every helix and cluster of helices, exactly as in Figure 1. The states that contribute at equilibrium, in addition to the fully folded state are state 46 (ABEGH) and state 62 (ABEFGH) with equilibrium probabilities of 0.023 and 0.457, respectively. State 46 lacks the BD pairing and the FH pairing,

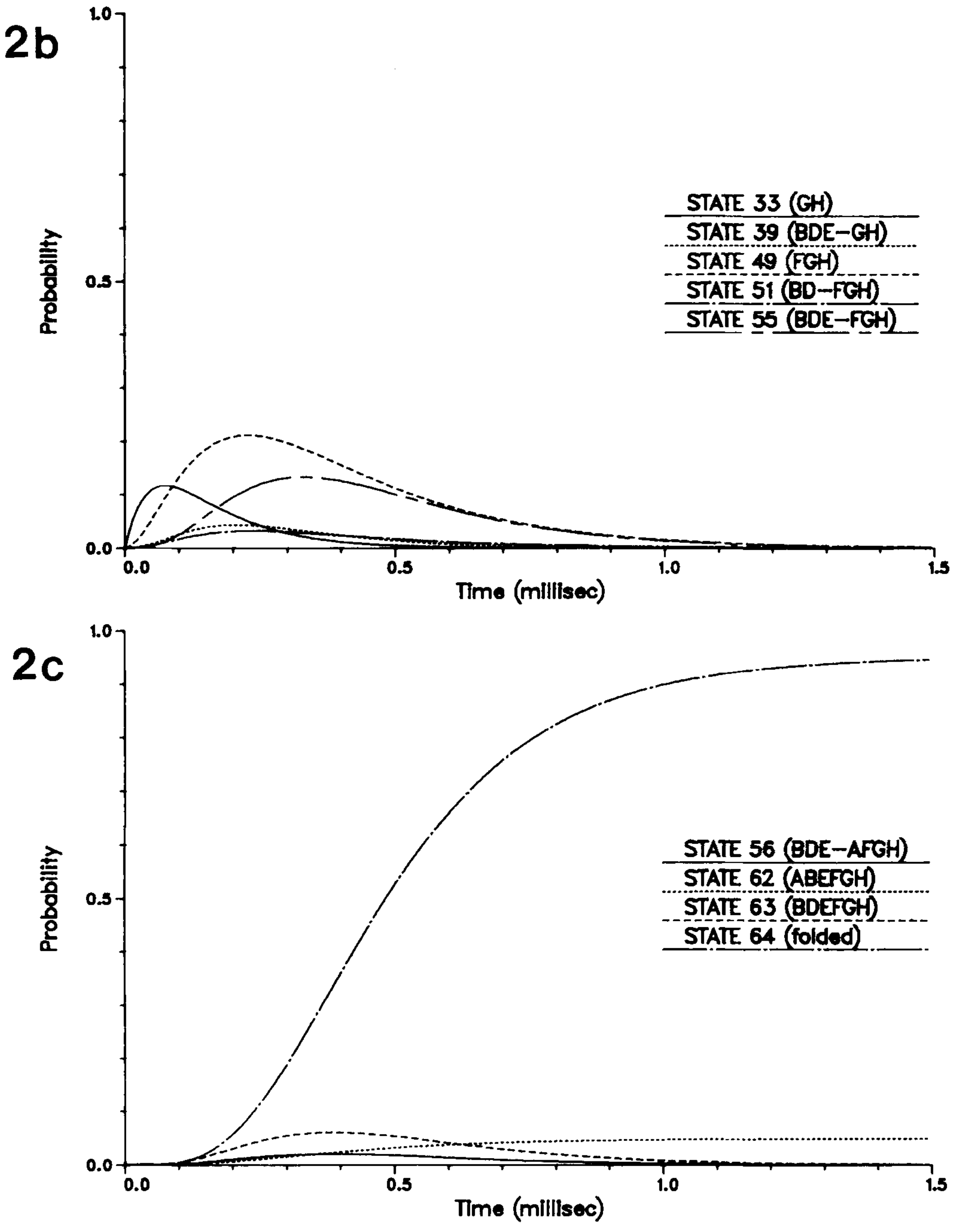
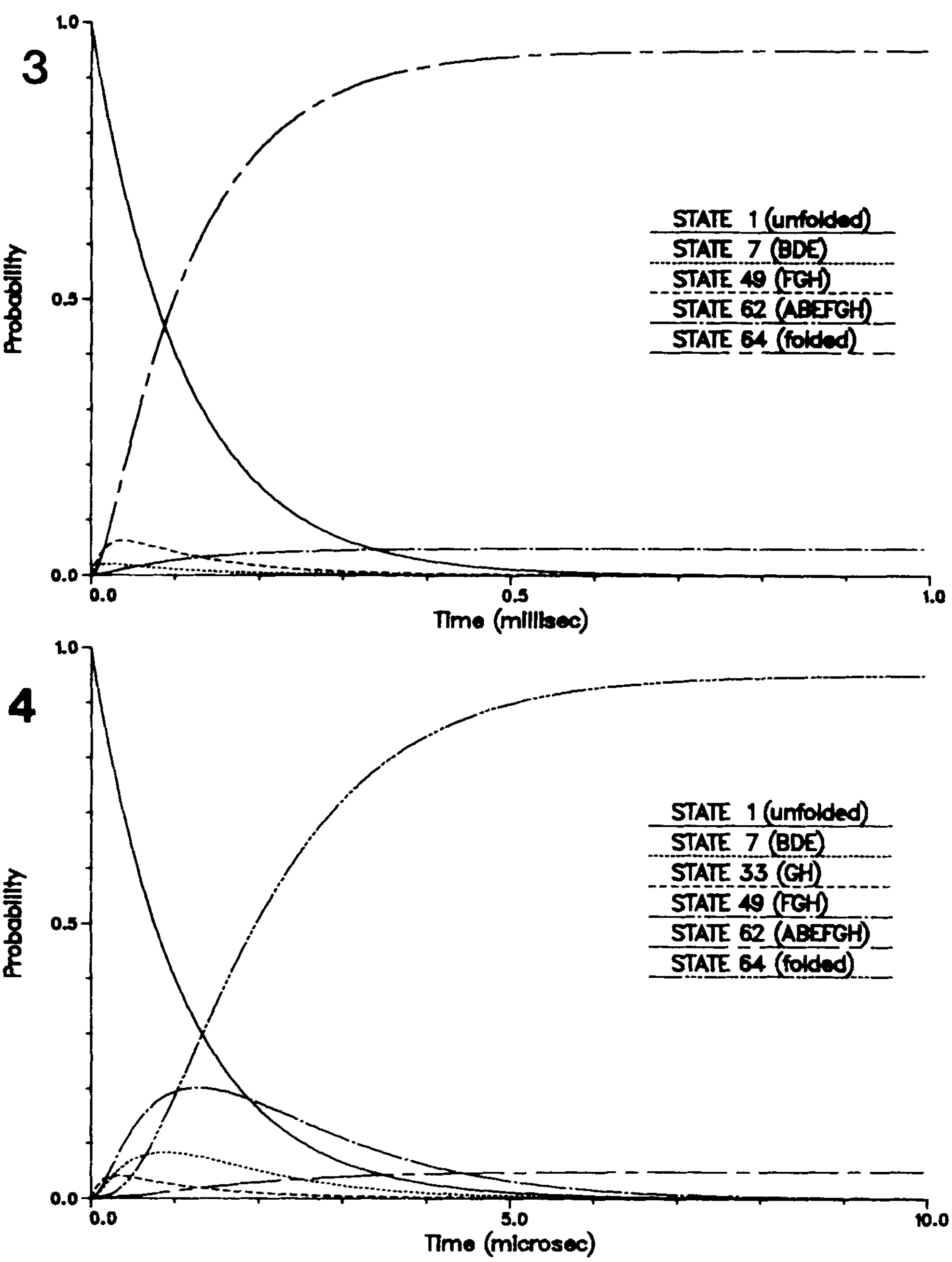


Fig. 2b–c

and state 62 lacks the BD pairing. Both of these states form with a lag period, as the most significant intermediate state 49 (FGH) begins to decay. Figure 6 shows a run with the same β values as used in Figure 4; at equilibrium state 62 (ABEFGH), the most stable intermediate, is present at 0.49, essentially equal in relative concentration to the fully folded state 64. In the early stages, state 49 (FGH) is dominant and state 7 (BDE) is present, but less so than under fully folding conditions. State 33 (GH) also contributes.

CONCLUSIONS

The diffusion-collision model is based on a physical picture that follows from the observed structural hierarchy of proteins and is consistent with what is known about the stability of protein structural elements or microdomains; i.e., secondary structures in isolation are generally unstable and must be stabilized by tertiary structural contacts. Based on the hypothesis that folding rates are governed by the diffusive encounter of microdomains, the diffusion-



Figs. 3 and 4

collision model allows calculations for the entire folding process. The calculations presented here show that with physically appropriate choices of parameters, the diffusion-collision model leads to multiple folding pathways and can explain the relatively short folding times that are observed experimentally. The most important intermediates are those formed by the coalescence of adjacent regions of the polypeptide chain, in accord with structural reasoning.[7] Thus, the present results support the suggestion of Harrison and Durbin[30] who have argued on evolutionary, structural, and experimental grounds that protein folding is most likely to proceed by multiple pathways in which native-like structural elements play a major role, and that a single unique path is not necessary to overcome the search problem. These ideas also figured prominently in the earlier paper by Cohen et al.[10]

By considering the significantly populated intermediates under fully folding conditions, one may formulate the folding diagram shown in Figure 7a. There are two main branches, one through the BDE cluster and the other through the FGH cluster. The branches converge at state 55 (BDE-FGH). The two clusters bond through the BG interaction to reach state 63 (BDEFGH), and then helix A associates with helix H

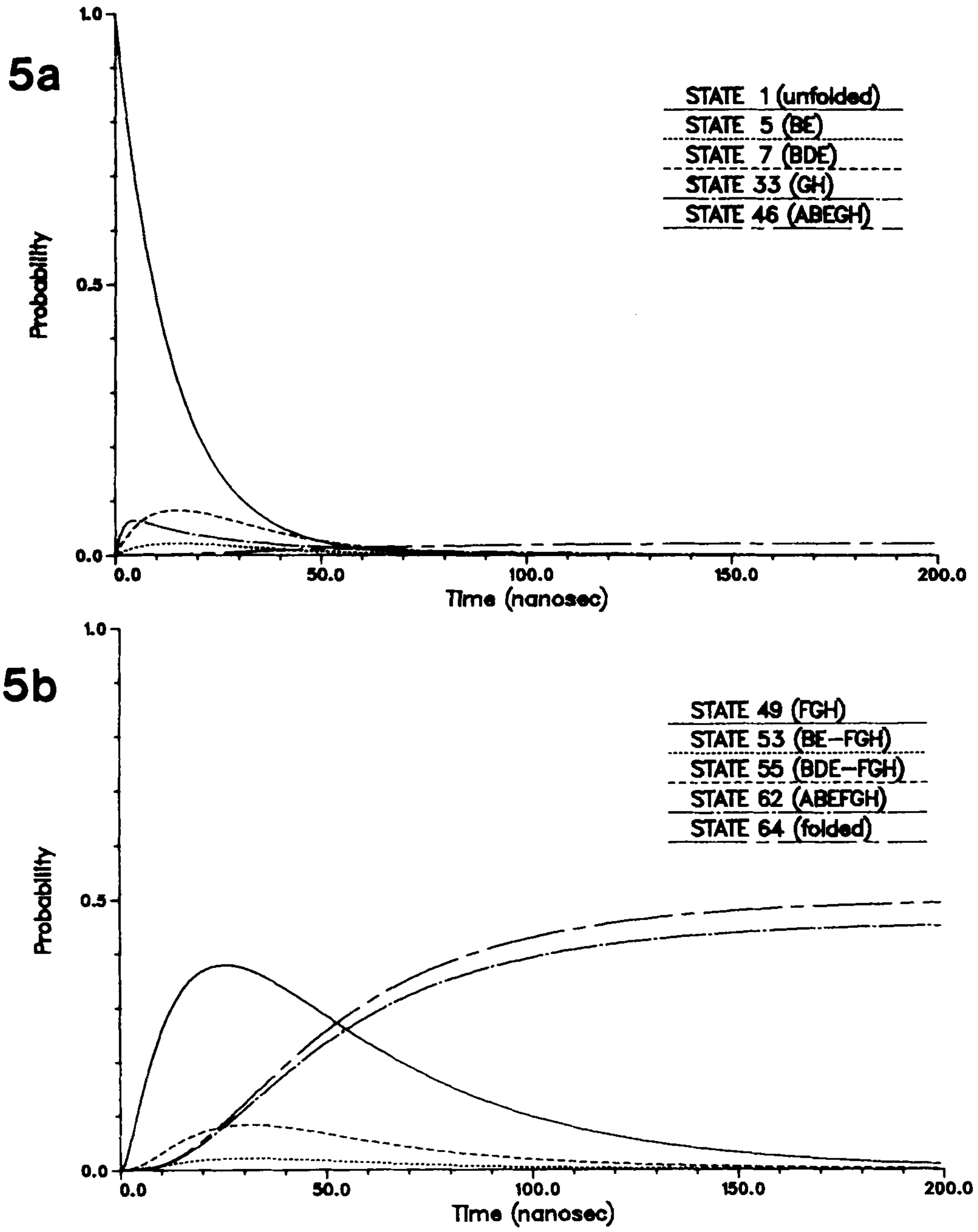


Fig. 5a-b

to produce the fully folded state 64. The fact that the same folding diagram is applicable for different coalescence parameter choices suggests that structural aspects (e.g., near neighbor contacts, magnitude of helix-helix interactions) dominate the folding. What the parameters change is the relative contributions of the intermediates and the overall time scale of the folding.

In experiments done under conditions that lead to partial folding, it might be assumed that the same process is followed as in complete folding and that the same intermediates contribute. Comparing Figures 1–4, where conditions strongly favor the folded state, with Figures 5 and 6, where the unfolding rate constants have been increased to mimic equilibrium conditions at the midpoint of folding-unfolding transitions, we see significant differences. Intermediates appear in Figures 5 and 6 that are not seen in Figures 1–4; and some of the species found in equilibrium at the midpoint were not seen as intermediates in the complete folding. These new states are displayed in terms of a pathway to incompletely folded states in Figure 7b. The possibility of different intermediates under different conditions arises because the diffusion-collision model allows for multiple "pathways," and the effect of environmental changes may not be the same for all of them. Thus, caution must be exercised in interpreting intermediates detected under

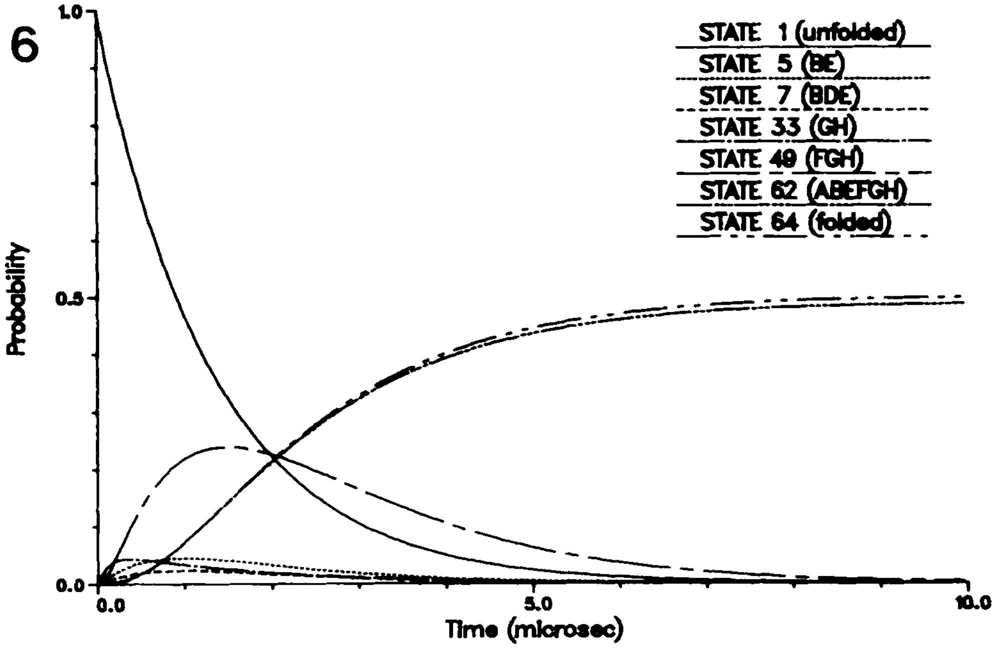


Fig. 6

experimental conditions that do not strongly favor folding: some of them may not be important in the actual folding process.

The particular shift of pathways seen here can be rationalized in structural terms. In the new final states, 48 and 62, the D and F helices, respectively, are not present in the cluster. Each of these helices is involved in only one of the six helix-helix bonds listed in Table I, and these are the two weakest bonds. Since the partially unfolding conditions are simulated by an increase in the dissociation rate, these two helices are the first ones affected. This is also true for the new intermediates shown in Figure 7b, all of which are missing the D and F helices.

We have not considered misfolded states here; this would require the enumeration and construction of non-native structures or segments of structure such as β-sheets or incorrectly packed helices and is beyond the scope of the present work. Studies of the pancreatic trypsin inhibitor[31] have shown that the formation of non-native disulfide bonds is an important part of the folding process and suggest that misfolded intermediate states should be included for a more complete and accurate model.

An important parameter in the model is β, the probability of finding a helix or helical cluster in a form that can coalesce on collision. Decreasing β increases the folding time and reduces the concentra-

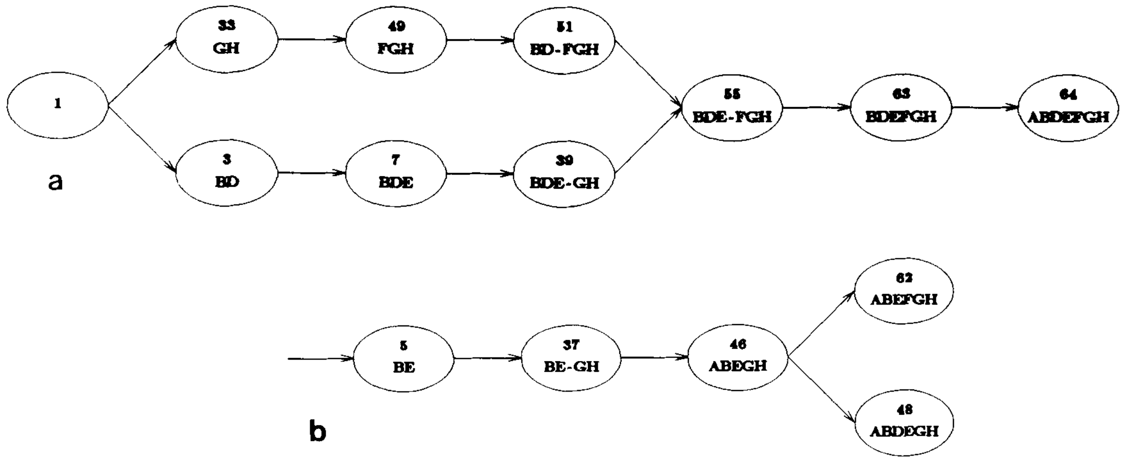


Fig. 7. **a**: The folding pathway implied by the major intermediates seen in runs 1–4 where solution conditions are presumed to be favorable to the native state. (Native state equilibrium probability = 0.95.) **b**: The additional folding pathway implied by the major intermediates seen in runs 5 and 6 where solution conditions are presumed to be in the unfolding transition zone. (Native state equilibrium probability = 0.50.) States 37 and 48 are shown here because they are on the pathway, but they do not appear in Figures 5 and 6 because they are too rapidly depleted to build up beyond the 0.02 plotting threshold.

TABLE V. Diffusion-Collision Model Runs for Apomyoglobin*

Run	One-helix microdomain (β)	Two-helix microdomain (β)	More-than-two-helix microdomain (β)	Frequency factor, θ (s^{-1})	Native equilibrium probability
1	1.0	1.0	1.0	2.94×10^{15}	0.95
2	0.01	0.01	0.01	2.17×10^{10}	0.95
3	0.01	1.0	1.0	2.19×10^{12}	0.95
4	0.1	0.5	1.0	2.39×10^{13}	0.95
5	1.0	1.0	1.0	5.36×10^{16}	0.50
6	0.1	0.5	1.0	4.50×10^{14}	0.50

*The parameters used in the model runs. The run numbers correspond to figure numbers showing the results in graphical form. β values are assigned to microdomains according to whether they contain one, two, or more helices. Products of these βs, as in Equation 8 are used in Equation 4 for the forward rates. The frequency factors, υ, have been adjusted as described earlier (see Dissociation Reaction) to yield the desired native equilibrium probability.

tion of intermediate states. This corresponds to the presence of a bottleneck in the folding. Low β values have the greatest effect in the early steps of folding where single helices rather than the more stable multihelix microdomains are involved. Transition rates into intermediate states are decreased more severely than subsequent folding transitions, so that the population of intermediates remains small. This suggests a difficulty for experiments attempting to detect folding intermediates: If the correct values for β are near 1, folding takes place on a submicrosecond time scale making stop-flow and temperature-jump experiments very difficult. On the other hand if β is small, the intermediates may not be present at detectable levels.

Recent studies on the helix-coil transition of the S peptide of ribonuclease[32,33] have shown that β can be highly variable and sequence specific. While we have made the simple assumption here that β is the same for all the myoglobin helices, it would be straightforward to implement sequence specificity as the data become available.

One question in protein folding concerns its dependence on the viscosity of the solvent. A viscosity dependence has been observed in the unfolding of ribonuclease A,[34] though other experiments suggested that folding is independent of viscosity.[35] In the present model, the viscosity dependence of the folding reaction depends on that of the switching parameter, $\lambda(\lambda = \lambda_1 + \lambda_2)$, which appears through α in Equation 6. If λ is inversely proportional to viscosity, then L is independent of viscosity and τ_b in Equation 4 is proportional to viscosity through the viscosity dependence of D. This type of behavior for λ is that found from Brownian dynamics calculations of helix-coil transitions[18] and collisional folding.[3] If λ is independent of viscosity, the viscosity dependence of L is determined by the magnitude of the αR terms in Equation 6. If $\alpha R \gg 1$ (switching is fast compared to diffusion across the space from R_{max} to R_{min}), the α term will dominate Equation 6, L will vary as the inverse square root of the viscosity, and the second term of Equation 4 for τ_b will vary as the square root of viscosity. If $\alpha R \ll 1$ (switching is slow compared to diffusion) then 1/L varies as α^2/R and the second term of Equation 4 is independent of viscosity. In most cases (unless β is very close to unity), the second term dominates in Equation 4, causing τ_b to be viscosity independent for $\alpha R < < 1$. Thus, it is possible for the rate of the elementary folding steps in the diffusion collision model to go from varying inversely with the viscosity to being independent of viscosity.

Physically, viscosity independent folding results when the details of the diffusion are unimportant; i.e., if switching of the microdomains to a favorable state for folding is the rate determining step. Although this does not apply for the range of parameter values used for myoglobin, where the main contribution to λ is the helix-coil transition rate, it is possible that there are cases in which folding rates are limited by much slower processes. One such case would be the cis-trans isomerization of proline, which appears to be the rate-limiting step in the slow refolding of several proteins.[1]

The diffusion-collision model in appropriate limits corresponds to some of the other folding models. If β is small for a single helix and increases for helix complexes, the first steps are slow and later steps are more rapid. This leads to kinetic behavior which corresponds to a nucleation model,[5] in which intermediates do not reach significant concentrations. The diffusion-collision model is also consistent with the growth-merge model[6] since adjacent microdomains tend to coalesce fastest. This is especially true if β is near one. In this limit the pair coalescence rates are dominated by the first term in Equation 4, which is most sensitive to the distance between microdomains.

Although the present analysis is based on a simplified model for protein folding, it does provide some new insights into the folding process. Refinements of the model are possible without introducing complexities that make the applications of the model too time-consuming. It is possible to include orientational effects explicitly in the model[12] and Brownian dynamics could be used to check the accuracy of the results.[3] It would be of interest to include coalescence to incorrect clusters and to consider explicitly the helix-coil transition. Also other types of microdomains (e.g., β strands) and more complex proteins (e.g., $\alpha\beta$ proteins) could be treated by the model.

Some recent experimental results[36–38] can be discussed in the context of the diffusion-collision model. All of the experiments suggest, in accord with the diffusion-collision model, that secondary structure formation occurs early in the folding process. Gilmanshin and Ptitsyn[36] have detected an intermediate in the folding of α-lactalbumin that forms within the 20 ms dead-time of the experiment and appears to have some nativelike secondary structure and buried aromatic side chains. Kuwajima et al.[37] have found that ferricytochrome c and β-lactoglobulin recover much of their secondary structure within 18 ms. Semisotnov et al.[38] have found that carbonic anhydrase recovers its secondary structure and forms a loosely packed hydrophobic core in an initial fast-folding process with a half-time of approximately 40 ms. The time scales of these fast-folding processes are somewhat slower than the slowest of our myoglobin simulations. This may be due to the arbitrary range of parameters (e.g., the values of β) used in the model, or possibly the neglect of misfolded states that could lead to slower folding. In all of these experimental studies, the initial fast folding, involving the generation of secondary structure, is followed by much slower (seconds to minutes) steps leading to full recovery of the native state. These slow steps may involve a variety of structural rearrangements, including proline isomerization, readjustment of the helix contacts, and loop reordering[38] that are not considered in the diffusion-collision model.

The picture of folding that the present work suggests is that rapid diffusion-collision folding proceeds, forming a substantial fraction of the secondary structure with loosely packed hydrophobic contacts, until a kinetic barrier, such as proline isomerizaton, is encountered and a slow step is required. If it is true that most of the general features of secondary structure and the overall fold are determined rapidly, it suggests that the daunting problem of obtaining a "low-resolution" nativelike structure is solved quite rapidly by nature.

NOTE ADDED IN PROOF

Wright et al.[39] have recently provided NMR measurements that support the diffusion-collision model.

REFERENCES

1. Kim, P.S., Baldwin, R.L. Specific intermediates in the folding of small proteins and mechanism of folding. AnnU. Rev. Biochem. 51:459–489, 1982.
2. Karplus, M., Weaver, D.L. Protein folding dynamics. Nature 260:404–406, 1976.
3. Lee, S., Karplus, M., Bashford, D., Weaver, D.L. Brownian dynamics simulation of protein folding: A study of the diffusion collision model. Biopolymers 26:481–506, 1987.
4. Karplus, M., Weaver, D.L. Diffusion-collision model for protein folding. Biopolymers 18:1421–1437, 1979.
5. Wetlaufer, D.B. Nucleation, rapid folding, and globular intrachain regions in proteins. Proc. Natl. Acad. Sci. USA 70:697–701, 1973.
6. Gō, N., Abe, H. Noninteracting local-structure model of folding and unfolding transitions in globular proteins. I. Formulation. Biopolymers 20:991–1011, 1981.
7. Levitt, M., Chothia, C. Structural patterns in globular proteins. Nature 261:552–558, 1976.
8. Lesk, A.M., Rose, G.D. Folding units in globular proteins. Proc. Natl. Acad. Sci. USA 78:4304–4308, 1981.
9. Ptitsyn, O.B., Rashin, A.A. A model of myoglobin self-organization. Biophys. Chem. 3:1–20, 1975.
10. Cohen, F.E., Sternberg, M.J.E., Phillips, D.C., Kuntz, I.D., Kollman, P.A. A diffusion-collision-adhesion model for the kinetics of myoglobin refolding. Nature 286:632–634, 1980.
11. Weaver, D.L. Alternative pathways in diffusion-collision controlled protein folding. Biopolymers 23:675–694, 1984.
12. Bashford, D. Fluctuation and rotation in diffusion-influenced systems. J. Chem. Phys. 85:6999–7010, 1986.
13. Gruenewald, G., Nicola, C.U., Lustig, A., Schwarz, G., Klump, H. Kinetics of the helix-coil transition of a polypeptide with nonionic side groups derived from ultrasonic relaxation measurements. Biophys. Chem. 9:137–147, 1979.
14. Hammes, G.G., Roberts, P.B. Dynamics of the helix-coil transition in poly-L-orthinine. J. Am. Chem. Soc. 91:1812–1816, 1969.
15. Zana, R. On the rate determining step for helix propagation in the helix-coil transition of polypeptides in solution. Biopolymers 14:2425–2428, 1975.
16. Inoue, S., Sano, T., Yakabe, Y., Ushio, H., Yasunaga, T. Kinetic studies of the helix coil transition in aqueous solutions of poly(-L-lysine). Biopolymers 18:681–691, 1979.
17. Bosterling, B., Engel, J. Kinetic studies on the helix-coil transition of fluorescent labeled poly(-L-lysine) by the temperature-jump technique. Biophys. Chem. 9:201–209, 1979.
18. McCammon, J.A., Northrup, S.H., Karplus, M., Levy, R.M. Helix-coil transitions in a simple polypeptide model. Biopolymers 19:2033–2045, 1980.
19. Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. J. Comp. Chem. 4:187–217, 1983.
20. Lee, B, Richards, F.M. The interpretation of Protein structures: estimation of static accessibility. J. Mol. Biol. 55:379–400, 1971.
21. Richmond, T.J., Richards, F.M. Packing of α-helices: Geometrical constraints and contact areas. J. Mol. Biol. 119:537–555, 1978.
22. Chothia, C. Hydrophobic bonding and accessible surface area in proteins. Nature 248:338–339, 1974.
23. Rossi Fanelli, A., Antonini, E., Caputo, A. Studies on the structure of hemoglobin. I. Physico-chemical properties of human globin. Biochim. Biophys. Acta 30:608–615, 1958.
24. Perutz, M.F., Kendrew, J.C., Watson, H.C. Structure and function of haemoglobin. II. Some relations between polypeptide chainconfiguration and amino acid sequence. J. Mol. Biol. 13:669–678,1965.
25. Lim, V.I., Ptitsyn, O.B. Constancy of the hydrophobic nucleus volume in myoglobin and hemoglobin molecules. Mol. Biol. (Mosk.) 4:372–382, 1970.
26. Lesk, A.M., Chothia, C. How different amino acid sequences determine similar protein structures: The structure and evolutionary dynamics of the globins. J. Mol. Biol. 136:225–270, 1980.
27. Sueki, M., Lee, S., Powers, S.P., Denton, J.B., Konishi, Y., Scheraga, H.A. Helix-coil stability constants for the naturally occurring amino acids in water. 22. Histidine parameters from random poly[(hydroxybutyl) glutamine-co-L-histidine]. Macromolecules 17:148–155, 1984.
28. Wagman, M.E., Dobson, C.M., Karplus, M. Proton NMR studies of theassociation and folding of glucagon in solution. FEBS Lett. 119:265–270, 1980.
29. Tanford, C. Protein denaturation. Adv. Protein Chem. 32:121–282, 1968.
30. Harrison, S.C., Durbin, R. Is there a simple pathway for the folding of a polypeptide chain? Proc. Natl. Acad. Sci. USA 82:4028–4030, 1985.
31. Creighton, T.E. Experimental studies of protein folding and unfolding. Prog. Biophys. Mol. Biol. 33:231–297, 1971.
32. Shoemaker, K.R., Kim, P.S., Brems, D.N., Marquese, S., York, E.J., Chaiken, I.M., Stewart, J.M., Baldwin, R.L. Nature of the charged group effect on the stability of the C-peptide helix. Proc. Natl. Acad. Sci. USA 82:2349–2353, 1985.

33. Mitchinson, C., Baldwin, R.L. The design and production of semisynthetic ribonucleases with increased thermostability by incorporation of S-peptide analogues with enhanced helical stability. Proteins 1:23–33, 1986.
34. Tsong, T.Y. Viscosity-dependent conformational relaxation of ribonuclease A in the thermal unfolding zone. Biochemistry 21:1493–1498, 1982.
35. Tsong, T.Y., Baldwin, R.L. Effects of solvent viscosity and different guanidine salts on the kinetics of ribonuclease A chain folding. Biopolymers 17:1669–1678, 1978.
36. Gilmanshin, R.I., Ptitsyn, O.B. An early intermediate of refolding α-lactalbumin forms within 20 ms. FEBS Lett. 223:327–329, 1987.
37. Kuwajima, K., Yamaya, H., Miwa, S., Sugai, S., Nagamura, T. Rapid formation of secondary structure framework in protein folding studied by stopped-flow circular dichroism. FEBS Lett. 221:115–118, 1987.
38. Semisotnov, G.V., Rodionova, N.A., Kutyshenko, V.P., Ebert, B., Blanck, J., Ptitsyn, O.B. Sequential mechanism of refolding of carbonic anhydrase B. FEBS Lett. 224:9–13, 1987.
39. Wright, P.E., Dyson, H.J., Lerner, R.A. Conformation of peptide fragments of proteins in aqueous solution: Implications for initiation of protein folding. Biochem. 27:7167–7175, 1988.
40. "IMSL Library Reference Manual." 9th Edition. IMSL Inc., 1982.

APPENDIX A: SOLUTION OF RATE EQUATIONS

We have the set of N first-order differential equation

$$\frac{dp_i(t)}{dt} = \sum_{j=1}^{N} R_{ij}\, p_j(t) \qquad (A1)$$

where $p_i(t)$ is the probability of finding the molecule in the i^{th} state at time t and R_{ij} is the transition probability per unit time for a molecule to go from the j^{th} state to the i^{th} state. It is convenient to define $y_i(t)$ by

$$p_i(t) \equiv y_i(t) + b_i \qquad (A2)$$

where b_i is the equilibrium probability defined by

$$\sum_{j=1}^{N} R_{ij} b_j = 0 \qquad (A3)$$

Then Equation A1 becomes

$$\frac{dy_i(t)}{dt} = \sum_{j=1}^{N} R_{ij} y_j(t) \qquad (A4)$$

Conservation of probability requires that

$$\sum_{i=1}^{N} \frac{dy_i}{dt} = \sum_{i=1}^{N}\sum_{j=1}^{N} R_{ij}p_j = \sum_{j=1}^{N} y_j \sum_{i=1}^{N} R_{ij} = 0 \qquad (A5)$$

Since this must hold for any set of y_i, the R matrix must satisfy the conditions

$$\sum_{i=1}^{N} R_{ij} = 0 \qquad (A6)$$

Conservation of probability also allows the elimination of one equation from the set (A1). Since

$$\sum_{i=1}^{N} p_i(t) = \sum_{i=1}^{N} b_i$$

Equation 12 requires that

$$y_N = -\sum_{i=1}^{N-1} y_i \qquad (A7)$$

Substitution into (A1) gives

$$\frac{dy_i}{dt} = \sum_{j=1}^{N-1} A_{ij} y_j \quad \text{where} \quad A_{ij} \equiv R_{ij} - R_{iN} \qquad (A8)$$

and A is an $(N-1) \times (N-1)$ matrix.

The solutions of Equation A8 are of the form

$$y_i = \sum_{k=1}^{N-1} C_{ik}\, e^{\lambda_k t} \qquad (A9)$$

Substitution of Equation A8 gives

$$\sum_{k=1}^{N-1} \left(\lambda_k C_{ik} - \sum_{j=1}^{N-1} A_{ij} C_{jk} \right) e^{\lambda_k t} = 0 \qquad (A10)$$

which requires that the term in brackets vanish for all i and k. This means that the columns of the matrix C are eigenvectors of matrix A and λ are the eigenvalues. However, the C matrix is still not completely determined because eigenvectors can always be multiplied by a constant; if Z is a matrix of eigenvectors we can write

$$C_{ik} = Z_{ik} f_k \qquad (A11)$$

and Equation A10 will be satisfied for any choice of f_k. The f_k are determined by the initial conditions

$$p_i(0) = b_i + \sum_{k=1}^{N-1} C_{ik} \qquad (A12)$$

from which

$$\sum_{k=1}^{N-1} Z_{ik} f_k = p_i\,(0) - b_i \qquad (A13)$$

The method of finding the equilibrium values b_i has to be discussed. Conservation of probability gives

$$b_N = 1 - \sum_{j=1}^{N-1} b_j \qquad (A14)$$

and substitution into Equation A13 results in

$$\sum_{j=1}^{N-1} A_{ij} b_j = -R_{iN} \qquad (A15)$$

Thus, solution of the rate equations reduces the solution of three linear algebra problems; namely, the eigenvalue problem (Eq. A10) and the linear equations (A15 and A13). This was accomplished by using subroutines from the IMSL library[40] on a VAX 11/780 computer; the subroutines EIGRF, LGINF, and LEQ2C were used to solve Equations A10, A15, and A13, respectively.