The early folding kinetics of apomyoglobin

ROHIT V. PAPPU^{1,2} AND DAVID L. WEAVER²

¹Department of Biochemistry & Molecular Biophysics, Washington University School of Medicine, St. Louis, Missouri 63110

²Molecular Modeling Laboratory, Department of Physics, Tufts University, Medford, Massachusetts 02155

(RECEIVED August 28, 1997; ACCEPTED October 16, 1997)

Abstract

The folding pathway of apomyoglobin has been experimentally shown to have early kinetic intermediates involving the A, B, G, and H helices. The earliest detected kinetic events occur on a ns to μ s time scale. We show that the early folding kinetics of apomyoglobin may be understood as the association of nascent helices through a network of diffusion-collision-coalescence steps $G + H \leftrightarrow GH + A \leftrightarrow AGH + B \leftrightarrow ABGH$ obtained by solving the diffusion-collision model in a chemical kinetics approximation. Our reproduction of the experimental results indicates that the model is a useful way to analyze folding data. One prediction from our fit is that the nascent A and H helices should be relatively more helix-like before coalescence than the other apomyoglobin helices.

Keywords: apomyoglobin; diffusion-collision model; folding kinetics; microdomains; nascent helices

Since the biological activity of a globular protein is determined by its native structure and internal flexibility (Brooks, et al, 1988), understanding how a protein folds and being able to produce an approximate three-dimensional structure is the culmination of the genetic code stored in the genes of living systems. To further this understanding, the kinetic aspects of the folding of apomyoglobin have been the subject of considerable experimental study (Hughson et al., 1990, 1991; Jennings & Wright, 1993; Shin et al., 1993a, 1993b; Waltho et al., 1993; Jennings et al., 1994; Eliezer et al., 1995; Loh et al., 1995; Ballew et al., 1996; Raymond et al., 1997). Elucidating the kinetic aspects of folding requires information on how the peptide elements that make up the protein interact dynamically to produce the native folding pattern. There is an apparent search problem in folding because the correct fold requires long-range interactions among the amino acids and statistical arguments (the Levinthal paradox (Levinthal, 1966, 1968)) show that a random search through all the possible conformations of even a moderately sized protein takes far too long for living systems. A possible mechanism by which a protein may make use of more sophisticated search procedures is the diffusion-collision model (Karplus & Weaver, 1976; 1994; Yapa & Weaver, 1996) in which fluctuating quasiparticles, called microdomains, which may be *incipient* secondary structures (α -helices, β -sheets) or hydrophobic clusters move via diffusion and collide with other microdomains. Coalescence may occur between *transiently* folded, correctly oriented microdomains, and folding would then proceed as a series of coalescence steps. Details of the kinetic pathway are determined by the physical and chemical properties of the microdomains and the barriers that need to be overcome for coalescence.

A diffusion-collision folding mechanism for a globular protein has a number of corollaries:

1. presence of microdomains;
2. nascent secondary structure forms (marginally and transiently) before tertiary structure;
3. kinetic intermediates accumulate transiently;
4. existence of folding pathways;
5. possible existence of incorrect (non-native) intermediates from non-native collisions;
6. existence of loosely folded multi-microdomain intermediate (s) (molten globule state);
7. solvent viscosity dependence of folding rates;
8. folding rates and favored pathway(s) dependent on properties of microdomains.

The available experimental and theoretical evidence for the diffusion-collision model has recently been reviewed (Karplus & Weaver, 1994) and supports many aspects of the model. In addition, more recent experiments have indicated the importance of diffusion-collision events in estimating the shortest time possible for a random coil protein to collapse to a compact structure (Hagen et al., 1996; McCammon, 1996); for the MLY mutant of the Arc repressor dimer to attain its folded structure (Baldwin, 1996; Waldburger et al., 1996) and for the solvent viscosity dependent kinetics of folding-unfolding of the cold-shock protein CspB (Jacob et al., 1997). A very recent simulation (Hoffmann & Knapp, 1997) of the folding of an helix-turn-helix motif, represented in atomic detail, is consistent with a number of the consequences of the diffusion-collision model.

Reprint requests to: Professor David Weaver, Molecular Modeling Laboratory, Department of Physics, Tufts University, Robinson Hall, Medford, MA 02155; e-mail: dweaver@tufts.edu.

Several of the consequences of the diffusion-collision model appear to occur in the apomyoglobin early folding kinetics studies (Hughson et al., 1990, 1991; Jennings & Wright, 1993; Shin et al., 1993a, 1993b; Waltho et al., 1993; Jennings et al., 1994; Eliezer et al., 1995; Loh et al., 1995; Ballew et al., 1996; Reymond et al., 1997). In order to make comparisons with the experimental kinetics results, we fit the previously formulated phenomenological chemical kinetics approximation to the diffusion-collision model (Bashford et al., 1988; Karplus & Weaver, 1994; Yapa & Weaver, 1996) to the apomyoglobin folding kinetic experiments. Used in this way, the diffusion-collision model is a useful tool for analysis of the kinetic data. By fitting the experimental results to the propensities of the nascent apomyoglobin helices to be partially folded and oriented upon collision with other helices (called beta values below), we obtain information on apomyoglobin helical folding probabilities. This allows us to estimate their relative stabilities during folding and is an important contribution of the kinetic results, since, at present, beta values for the folding of a particular protein must be obtained from experiment. In what follows, we briefly describe the calculational details and apply the model to the early folding kinetics of apomyoglobin, taking into account the recent experimental evidence.

Results

Apomyoglobin folding model

Apomyoglobin (Bernstein et al., 1977) has eight helices labeled A $\rightarrow$ H. The eight helices are involved in 13 helix-helix hydrophobic interactions ($>250 \text{ \AA}^2$), as measured by solvent accessible area loss (Lee & Richards, 1971; Weaver, 1992) upon folding in the native state. The six largest interactions are BG, GH, BE, FH, AH, and AE (in decreasing order of area loss). Therefore, in a first-approximation to the folding kinetics as calculated from the diffusion-collision model, we neglect the two small helices C and D and consider the protein as composed of A, B, E, F, G, and H, this group containing the six largest pairwise interactions. With six pairings, there are 64 states of native microdomain aggregates for the folding apomyoglobin, including the initial, completely unfolded state (state 1) and the completely folded state (state 64). Figure 1 shows the 192 possible transitions and 720 pathways between the completely unfolded (none of the 6 helix-helix contacts made) state 1 and state 64, which has all 6 helix-helix contacts made. Figure 1 illustrates the complexity of folding pathways obtained from all the combinatorial possibilities of the order in which helix-helix contacts are made. As the calculations below show, not all of the possible pathways are equally likely.

The folding kinetics of apomyoglobin are taken to be transitions among the 64 states which occur with the formation or breaking of only one helix-helix contact (Bashford et al., 1988). These transitions are described in the first column of Table 1 (e.g., $2 \rightarrow 6$ means the transition between state 2 with the AH pair and state 6 with the AH and BE pairs). Table 1 gives the microdomains in the initial state, the helix-helix interaction causing the transition, the final state microdomains and the pair of microdomains that coalesce in the folding reaction (this pair would dissociate in the corresponding unfolding process). Also in Table 1 are the descriptions of the relative diffusion coefficient, the minimum distance between microdomains ($R_{\min}$, the inner boundary for diffusion) which is the spherical shell of closest approach (governed by the van der Waals envelopes of the microdomains) and the maximum

Fig. 1. Schematic diagram of the possible native folding pathways and transitions for the six microdomain approximation to apomyoglobin. The transitions are quantified in Table 1. The numerical labels correspond to the transition states listed in Table 1.

distance between microdomains ($R_{\max}$). This is the outer diffusion boundary, the maximum radial separation of the microdomains determined by the intervening microdomains and the polypeptide chain between microdomains $R_{\min}$ and $R_{\max}$ are illustrated in Figure 2A. We take as an estimate of the folding time

$$\tau_f = \frac{l^2}{D} + \frac{L\Delta V(1 - \beta)}{\beta DA} \quad (1)$$

as described previously (Bashford et al., 1988; Karplus & Weaver, 1994; Yapa & Weaver, 1996). The volume available for diffusion of each microdomain pair ΔV , their relative target surface area for collisions A , their relative diffusion coefficient D and their relative geometry parameter l^2 are calculated for diffusion in a spherical space. The parameter l^2 has units of length squared and the value

$$l^2 = \frac{R_{\max}^2}{3} \frac{\left(1 - \frac{9}{5}\epsilon + \epsilon^2 - \frac{1}{5}\epsilon^6\right)}{\epsilon(1 - \epsilon^3)} \quad (2)$$

where $\epsilon = R_{\min}/R_{\max}$. The diffusion volume (denoted by ΔV) is $\frac{3}{4}\tau(R_{\max}^3 - R_{\min}^3)$. The parameter L has units of length. Its value is

$$\frac{1}{L} = \frac{1}{R_{\min}} + \alpha \frac{\alpha R_{\max} \tanh[\alpha(R_{\max} - R_{\min})] - 1}{\alpha R_{\max} - \tanh[\alpha(R_{\max} - R_{\min})]} \quad (3)$$

where

$$\alpha \equiv \left(\frac{\lambda_1 + \lambda_2}{D}\right)^{1/2}. \quad (4)$$

Table 1. Apomyoglobin transitions: States, bonds, and parameters

Transition	Initial state MDs	Bond formed	Final state MDs	MDs coalescing	D	$R_{\min}$	$R_{\max}$
1 $\rightarrow$ 2	A-B-E-F-G-H	AH	B-E-F-G-AH	A + H	$D_A + D_H$	$r_A + r_H$	$r_A + c_{AB} + 2r_B + c_{BE} + 2r_E + c_{EF} + 2r_F + c_{FG} + 2r_G + c_{GH} + r_H$
1 $\rightarrow$ 3	A-B-E-F-G-H	AE	B-F-G-H-AE	A + E	$D_A + D_E$	$r_A + r_E$	$r_A c_{AB} + 2r_B + c_{BE} + r_E$
1 $\rightarrow$ 5	A-B-E-F-G-H	BE	A-F-G-H-BE	B + E	$D_B + D_E$	$r_B + r_E$	$r_B + c_{BE} + r_E$
1 $\rightarrow$ 9	A-B-E-F-G-H	BG	A-E-F-H-BG	B + G	$D_B + D_G$	$r_B + r_G$	$r_B + c_{BE} + 2r_E + c_{EF} + 2r_F + c_{FG} + r_G$
1 $\rightarrow$ 17	A-B-E-F-G-H	FH	A-B-E-G-FH	F + H	$D_F + D_H$	$r_F + r_H$	$r_F + c_{FG} + 2r_G + c_{GH} + r_H$
1 $\rightarrow$ 33	A-B-E-F-G-H	GH	A-B-E-F-GH	G + H	$D_G + D_H$	$r_G + r_H$	$r_G + c_{GH} + r_H$
2 $\rightarrow$ 4	B-E-F-G-AH	AE	B-F-G-AEH	AH + E	$D_{AH} + D_E$	$r_{AH} + r_E$	$r_{AH} + c_{GH} + 2r_G + c_{FG} + 2r_F + c_{EF} + r_E$
2 $\rightarrow$ 6	B-E-F-G-AH	BE	F-G-AH-BE	B + E	$D_B + D_E$	$r_B + r_E$	$r_B + c_{BE} + r_E$
2 $\rightarrow$ 10	B-E-F-G-AH	BG	E-F-AH-BG	B + G	$D_B + D_G$	$r_B + r_G$	$r_B + c_{AB} + 2r_{AH} + c_{GH} + r_G$
2 $\rightarrow$ 18	B-E-F-G-AH	FH	B-E-G-AFH	AH + F	$D_{AH} + D_F$	$r_{AH} + r_F$	$r_{AH} + c_{GH} + 2r_G + c_{FG} + r_F$
2 $\rightarrow$ 34	B-E-F-G-AH	GH	B-E-F-AGH	AH + G	$D_{AH} + D_G$	$r_{AH} + r_G$	$r_{AH} + c_{GH} + r_G$
3 $\rightarrow$ 4	B-F-G-H-AE	AH	B-F-G-AEH	AE + H	$D_{AE} + D_H$	$r_{AE} + r_H$	$r_{AE} + c_{EF} + 2r_F + c_{FG} + 2r_G + c_{GH} + r_H$
3 $\rightarrow$ 7	B-F-G-H-AE	BE	F-G-H-AE	AE + B	$D_{AE} + D_B$	$r_{AE} + r_B$	$r_{AE} + c_{AB} + r_B$
3 $\rightarrow$ 11	B-F-G-H-AE	BG	F-H-AE-BG	B + G	$D_B + D_G$	$r_B + r_G$	$r_B + c_{AB} + 2r_{AE} + c_{EF} + 2r_F + c_{FG} + r_G$
3 $\rightarrow$ 19	B-F-G-H-AE	FH	B-G-AE-FH	F + H	$D_F + D_H$	$r_F + r_H$	$r_F + c_{FG} + 2r_G + c_{GH} + r_H$
3 $\rightarrow$ 35	B-F-G-H-AE	GH	B-F-AE-GH	G + H	$D_G + D_H$	$r_G + r_H$	$r_G + c_{GH} + r_H$
5 $\rightarrow$ 6	A-F-G-H-BE	AH	F-H-AH-BE	A + H	$D_A + D_H$	$r_A + r_H$	$r_A + c_{AB} + 2r_{BE} + c_{EF} + 2r_F + c_{FG} + 2r_G + c_{GH} + r_H$
5 $\rightarrow$ 7	A-F-G-H-BE	AE	F-G-H-AE	A + BE	$D_A + D_{BE}$	$r_A + r_{BE}$	$r_A + c_{AB} + r_{BE}$
5 $\rightarrow$ 13	A-F-G-H-BE	BG	A-F-H-BEG	BE + G	$D_{BE} + D_G$	$r_{BE} + r_G$	$r_{BE} + c_{EF} + 2r_F + c_{FG} + r_G$
5 $\rightarrow$ 21	A-F-G-H-BE	FH	A-G-BE-FH	F + H	$D_F + D_H$	$r_F + r_H$	$r_F + c_{FG} + 2r_G + c_{GH} + r_H$
5 $\rightarrow$ 37	A-F-G-H-BE	GH	A-F-BE-GH	G + H	$D_G + D_H$	$r_G + r_H$	$r_G + c_{GH} + r_H$
9 $\rightarrow$ 10	A-E-F-H-BG	AH	E-F-AH-BG	A + H	$D_A + D_H$	$r_A + r_H$	$r_A + c_{AB} + 2r_{BG} + c_{GH} + r_H$
9 $\rightarrow$ 11	A-E-F-H-BG	AE	F-H-AE-BG	A + E	$D_A + D_E$	$r_A + r_E$	$r_A + c_{AB} + 2r_{BG} + c_{FG} + 2r_F + c_{EF} + r_E$
9 $\rightarrow$ 13	A-E-F-H-BG	BE	A-F-H-BEG	BG + E	$D_{BG} + D_E$	$r_{BG} + r_E$	$r_{BG} + c_{FG} + 2r_F + c_{EF} + r_E$
9 $\rightarrow$ 25	A-E-F-H-BG	FH	A-E-BG-FH	F + H	$D_F + D_H$	$r_F + r_H$	$r_F + c_{FG} + 2r_{BG} + c_{GH} + r_H$
9 $\rightarrow$ 41	A-E-F-H-BG	GH	A-E-F-BGH	BG + H	$D_{BG} + D_H$	$r_{BG} + r_H$	$r_{BG} + c_{GH} + r_H$
17 $\rightarrow$ 18	A-B-E-G-FH	AH	B-E-G-AFH	A + FH	$D_A + D_{FH}$	$r_A + r_{FH}$	$r_A + c_{AB} + 2r_B + c_{BE} + 2r_E + c_{EF} + r_{FH}$
17 $\rightarrow$ 19	A-B-E-G-FH	AE	B-G-AE-FH	A + E	$D_A + D_E$	$r_A + r_E$	$r_A + c_{AB} + 2r_B + c_{BE} + r_E$
17 $\rightarrow$ 21	A-B-E-G-FH	BE	A-G-BE-FH	B + E	$D_B + D_E$	$r_B + r_E$	$r_B + c_{BE} + r_E$
17 $\rightarrow$ 25	A-B-E-G-FH	BG	A-E-BG-FH	B + G	$D_B + D_G$	$r_B + r_G$	$r_B + c_{BE} + 2r_E + c_{EF} + 2r_{FH} + c_{FG} + r_G$
17 $\rightarrow$ 49	A-B-E-G-FH	GH	A-B-E-FGH	FH + G	$D_{FH} + D_G$	$r_{FH} + r_G$	$r_G + c_{FG} + r_{FH}$
33 $\rightarrow$ 34	A-B-E-F-GH	AH	B-E-F-AGH	A + GH	$D_A + D_{GH}$	$r_A + r_{GH}$	$r_A + c_{AB} + 2r_B + c_{BE} + 2r_E + c_{EF} + 2r_F + c_{FG} + r_{GH}$
33 $\rightarrow$ 35	A-B-E-F-GH	AE	B-F-AE-GH	A + E	$D_A + D_E$	$r_A + r_E$	$r_A + c_{AB} + 2r_B + c_{BE} + r_E$
33 $\rightarrow$ 37	A-B-E-F-GH	BE	A-F-BE-GH	B + E	$D_B + D_E$	$r_B + r_E$	$r_B + c_{BE} + r_E$
33 $\rightarrow$ 41	A-B-E-F-GH	BG	A-E-F-BGH	B + GH	$D_B + D_{GH}$	$r_B + r_{GH}$	$r_B + c_{BE} + 2r_E + c_{EF} + 2r_F + c_{FG} + r_{GH}$
33 $\rightarrow$ 49	A-B-E-F-GH	FH	A-B-E-FGH	F + GH	$D_F + D_{GH}$	$r_F + r_{GH}$	$r_F + c_{FG} + r_{GH}$
4 $\rightarrow$ 8	B-F-G-AEH	BE	F-G-AE	AEH + B	$D_{AEH} + D_B$	$r_{AEH} + r_B$	$r_B + c_{AB} + r_{AEH}$
4 $\rightarrow$ 12	B-F-G-AEH	BG	F-BG-AEH	B + G	$D_B + D_G$	$r_B + r_G$	$r_B + c_{AB} + 2r_{AEH} + c_{GH} + r_G$
4 $\rightarrow$ 20	B-F-G-AEH	FH	B-G-AEFH	AEH + F	$D_{AEH} + D_F$	$r_{AEH} + r_F$	$r_F + c_{EF} + r_{AEH}$
4 $\rightarrow$ 36	B-F-G-AEH	GH	B-F-AEGH	AEH + G	$D_{AEH} + D_G$	$r_{AEH} + r_G$	$r_G + c_{GH} + r_{AEH}$
6 $\rightarrow$ 8	F-G-AH-BE	AE	F-G-AE	AH + BE	$D_{AH} + D_{BE}$	$r_{AH} + r_{BE}$	$r_{AH} = c_{AB} + r_{BE}$
6 $\rightarrow$ 14	F-G-AH-BE	BG	F-AH-BEG	BE + G	$D_{BE} + D_G$	$r_{BE} + r_G$	$r_{BE} + c_{AB} + 2r_{AH} + c_{GH} + r_G$
6 $\rightarrow$ 22	F-G-AH-BE	FH	G-BE-AFH	AH + F	$D_{AH} + D_F$	$r_{AH} + r_F$	$r_F + c_{EF} + 2r_{BE} + c_{AB} + r_{AH}$
6 $\rightarrow$ 38	F-G-AH-BE	GH	F-BE-AGH	AH + G	$D_{AH} + D_G$	$r_{AH} + r_G$	$r_G + c_{GH} + r_{AH}$
7 $\rightarrow$ 8	F-G-H-AE	AH	F-G-AE	ABE + H	$D_{ABE} + D_H$	$r_{ABE} + r_H$	$r_{ABE} + c_{EF} + 2r_F + c_{FG} + 2r_G + c_{GH} + r_H$
7 $\rightarrow$ 15	F-G-H-AE	BG	F-H-ABEG	ABE + G	$D_{ABE} + D_G$	$r_{ABE} + r_G$	$r_{ABE} + c_{EF} + 2r_F + c_{FG} + r_G$
7 $\rightarrow$ 23	F-G-H-AE	FH	G-ABE-FH	F + H	$D_F + D_H$	$r_F + r_H$	$r_F + c_{FG} + 2r_G + c_{GH} + r_H$
7 $\rightarrow$ 39	F-G-H-AE	GH	F-ABE-GH	G + H	$D_G + D_H$	$r_G + r_H$	$r_G + c_{GH} + r_H$
10 $\rightarrow$ 12	E-F-AH-BG	AE	F-BG-AEH	AH + E	$D_{AH} + D_E$	$r_{AH} + r_E$	$r_{AH} + c_{AB} + 2r_{BG} + c_{FG} + 2r_F + c_{EF} + r_E$
10 $\rightarrow$ 14	E-F-AH-BG	BE	F-AH-BEG	BG + E	$D_{BG} + D_E$	$r_{BG} + r_E$	$r_{BG} + c_{FG} + 2r_F + c_{EF} + r_E$
10 $\rightarrow$ 26	E-F-AH-BG	FH	E-BG-AFH	AH + F	$D_{AH} + D_F$	$r_{AH} + r_F$	$r_F + c_{FG} + 2r_{BG} + c_{AB} + r_{AH}$
10 $\rightarrow$ 42	E-F-AH-BG	GH	E-F-ABGH	AH + BG	$D_{AH} + D_{BG}$	$r_{AH} + r_{BG}$	$r_{BG} + c_{AB} + r_{AH}$
11 $\rightarrow$ 12	F-H-AE-BG	AH	F-BG-AEH	AE + H	$D_{AE} + D_H$	$r_{AE} + r_H$	$r_{AE} + c_{AB} + 2r_{BG} + c_{GH} + r_H$
11 $\rightarrow$ 15	F-H-AE-BG	BE	F-H-ABEG	AE + BG	$D_{AE} + D_{BG}$	$r_{AE} + r_{BG}$	$r_{AE} + c_{AB} + r_{BG}$
11 $\rightarrow$ 27	F-H-AE-BG	FH	AE-BG-FH	F + H	$D_F + D_H$	$r_F + r_H$	$r_F + c_{FG} + 2r_{BG} + c_{GH} + r_H$
11 $\rightarrow$ 43	F-H-AE-BG	GH	F-AE-BGH	BG + H	$D_{BG} + D_H$	$r_{BG} + r_H$	$r_{BG} + c_{GH} + r_H$
13 $\rightarrow$ 14	A-F-H-BEG	AH	F-AH-BEG	A + H	$D_A + D_H$	$r_A + r_H$	$r_A + c_{AB} + 2r_{BEG} + c_{GH} + r_H$
13 $\rightarrow$ 15	A-F-H-BEG	AE	F-H-ABEG	A + BEG	$D_A + D_{BEG}$	$r_A + r_{BEG}$	$r_A + c_{AB} + r_{BEG}$

(continued)

Table 1. *Continued*

Transition	Initial state MDs	Bond formed	Final state MDs	MDs coalescing	D	$R_{\min}$	$R_{\max}$
13 → 29	A-F-H-BEG	FH	A-FH-BEG	F + H	$D_F + D_H$	$r_F + r_H$	$rF + cFG + 2rBEG + cGH + rH$
13 → 45	A-F-H-BEG	GH	A-F-BEGH	BEG + H	$D_{BEG} + D_H$	$r_{BEG} + r_H$	$rBEG + cGH + rH$
18 → 20	B-E-G-AFH	AE	B-G-AEFH	AFH + E	$D_{AFH} + D_E$	$r_{AFH} + r_E$	$rAFH + cEF + rE$
18 → 22	B-E-G-AFH	BE	G-BE-AFH	B + E	$D_B + D_E$	$r_B + r_E$	$rB + cAB + 2rAFH + cEF + rE$
18 → 26	B-E-G-AFH	BG	E-BG-AFH	B + G	$D_B + D_G$	$r_B + r_G$	$rB + cAB + 2rAFH + cFG + rG$
18 → 50	B-E-G-AFH	GH	B-E-AFGH	AFH + G	$D_{AFH} + D_G$	$r_{AFH} + r_G$	$rAFH + cFG + rG$
19 → 20	B-G-AE-FH	AH	B-G-AEFH	AE + FH	$D_{AE} + D_{FH}$	$r_{AE} + r_{FH}$	$rAE + cEF + rFH$
19 → 23	B-G-AE-FH	BE	G-ABE-FH	AE + B	$D_{AE} + D_B$	$r_{AE} + r_B$	$rB + cAB + rAE$
19 → 27	B-G-AE-FH	BG	AE-BG-FH	B + G	$D_B + D_G$	$r_B + r_G$	$rB + cAB + 2rAE + cEF + 2rFH$ $+ cFG + rG$
19 → 51	B-G-AE-FH	GH	B-AE-FGH	FH + G	$D_{FH} + D_G$	$r_{FH} + r_G$	$rG + cFG + rFH$
21 → 22	A-G-BE-FH	AH	G-BE-AFH	A + FH	$D_A + D_{FH}$	$r_A + r_{FH}$	$rA + cAB + 2rBE + cEF + rFH$
21 → 23	A-G-BE-FH	AE	G-ABE-FH	A + BE	$D_A + D_{BE}$	$r_A + r_{BE}$	$rA + cAB + rBE$
21 → 29	A-G-BE-FH	BG	A-FH-BEG	BE + G	$D_{BE} + D_G$	$r_{BE} + r_G$	$rBE + cEF + 2rFH + cFG + rG$
21 → 53	A-G-BE-FH	GH	A-BE-FGH	FH + G	$D_{FH} + D_G$	$r_{FH} + r_G$	$rG + cFG + rFH$
25 → 26	A-E-BG-FH	AH	E-BG-AFH	A + FH	$D_A + D_{FH}$	$r_A + r_{FH}$	$rA + cAB + 2rBG + cFG + rFH$
25 → 27	A-E-BG-FH	AE	AE-BG-FH	A + E	$D_A + D_E$	$r_A + r_E$	$rA + cAB + 2rBG + cFG + 2rFH$ $+ cEF + rE$
25 → 29	A-E-BG-FH	BE	A-FH-BEG	BG + E	$D_{BG} + D_E$	$r_{BG} + r_E$	$rBG + cFG + 2rFH + cEF + rE$
25 → 57	A-E-BG-FH	GH	A-E-BFGH	BG + FH	$D_{BG} + D_{FH}$	$r_{BG} + r_{FH}$	$rBG + cFG + rFH$
34 → 36	B-E-F-AGH	AE	B-F-AEGH	AGH + E	$D_{AGH} + D_E$	$r_{AGH} + r_E$	$rAGH + cFG + 2rF + cEF + rE$
34 → 38	B-E-F-AGH	BE	F-BE-AGH	B + E	$D_B + D_E$	$r_B + r_E$	$rB + cBE + rE$
34 → 42	B-E-F-AGH	BG	E-F-ABGH	AGH + B	$D_{AGH} + D_B$	$r_{AGH} + r_B$	$rB + cAB + rAGH$
34 → 50	B-E-F-AGH	FH	B-E-AFGH	AGH + F	$D_{AGH} + D_F$	$r_{AGH} + r_F$	$rF + cFG + rAGH$
35 → 36	B-F-AE-GH	AH	B-F-AEGH	AE + GH	$D_{AE} + D_{GH}$	$r_{AE} + r_{GH}$	$rAE + cEF + 2rF + cFG + rGH$
35 → 39	B-F-AE-GH	BE	F-GH-ABE	AE + B	$D_{AE} + D_B$	$r_{AE} + r_B$	$rAE + cAB + rB$
35 → 43	B-F-AE-GH	BG	F-AE-BGH	B + GH	$D_B + D_{GH}$	$r_B + r_{GH}$	$rB + cAB + 2rAE + cEF + 2rF$ $+ cFG + rGH$
35 → 51	B-F-AE-GH	FH	B-AE-FGH	F + GH	$D_F + D_{GH}$	$r_F + r_{GH}$	$rF + cFG + rGH$
37 → 38	A-F-BE-GH	AH	F-BE-AGH	A + GH	$D_A + D_{GH}$	$r_A + r_{GH}$	$rA + cAB + 2rBE + cEF + 2rF$ $+ cFG + rGH$
37 → 39	A-F-BE-GH	AE	F-GH-ABE	A + BE	$D_A + D_{BE}$	$r_A + r_{BE}$	$rA + cAB + rBE$
37 → 45	A-F-BE-GH	BG	A-F-BEGH	BE + GH	$D_{BE} + D_{GH}$	$r_{BE} + r_{GH}$	$rBE + cEF + 2rF + cFG + rGH$
37 → 53	A-F-BE-GH	FH	A-BE-FGH	F + GH	$D_F + D_{GH}$	$r_F + r_{GH}$	$rF + cFG + rGH$
41 → 42	A-E-F-BGH	AH	E-F-ABGH	A + BGH	$D_A + D_{BGH}$	$r_A + r_{BGH}$	$rA + cAB + rBGH$
41 → 43	A-E-F-BGH	AE	F-AE-BGH	A + E	$D_A + D_E$	$r_A + r_E$	$rA + cAB + 2rBGH + cFG + 2rF$ $+ cEF + rE$
41 → 45	A-E-F-BGH	BE	A-F-BEGH	BGH + E	$D_{BGH} + D_E$	$r_{BGH} + r_E$	$rBGH + cFG + 2rF + cEF + rE$
41 → 57	A-E-F-BGH	FH	A-E-BFGH	BGH + F	$D_{BGH} + D_F$	$r_{BGH} + r_F$	$rF + cFG + rBGH$
49 → 50	A-B-E-FGH	AH	B-E-AFGH	A + FGH	$D_A + D_{FGH}$	$r_A + r_{FGH}$	$rA + cAB + 2rB + cBE + 2rE$ $+ cEF + rFGH$
49 → 51	A-B-E-FGH	AE	B-AE-FGH	A + E	$D_A + D_E$	$r_A + r_E$	$rA + cAB + 2rB + cBE + rE$
49 → 53	A-B-E-FGH	BE	A-BE-FGH	B + E	$D_B + D_E$	$r_B + r_E$	$rB + cBE + rE$
49 → 57	A-B-E-FGH	BG	A-E-BFGH	B + FGH	$D_B + D_{FGH}$	$r_B + r_{FGH}$	$rB + cBE + 2rE + cEF + rFGH$
8 → 16	F-G-ABEH	BG	F-ABEGH4	ABEH + G	$D_{ABEH} + D_G$	$r_{ABEH} + r_G$	$rABEH + cGH + rG$
8 → 24	F-G-ABEH	FH	G-ABEFH	ABEH + F	$D_{ABEH} + D_F$	$r_{ABEH} + r_F$	$rF + cEF + rABEH$
8 → 40	F-G-ABEH	GH	F-ABEGH4	ABEH + G	$D_{ABEH} + D_G$	$r_{ABEH} + r_G$	$rG + cGH + rABEH$
12 → 16	F-BG-AEH	BE	F-ABEGH4	AEH + BG	$D_{AEH} + D_{BG}$	$r_{AEH} + r_{BG}$	$rBG + cAB + rAEH$
12 → 28	F-BG-AEH	FH	BG-AEFH	AEH + F	$D_{AEH} + D_F$	$r_{AEH} + r_F$	$rF + cEF + rAEH$
12 → 44	F-BG-AEH	GH	F-ABEGH4	AEH + BG	$D_{AEH} + D_{BG}$	$r_{AEH} + r_{BG}$	$rBG + cAB + rAEH$
14 → 16	F-AH-BEG	AE	F-ABEGH4	AH + BEG	$D_{AH} + D_{BEG}$	$r_{AH} + r_{BEG}$	$rAH + cAB + rBEG$
14 → 30	F-AH-BEG	FH	AFH-BEG	AH + F	$D_{AH} + D_F$	$r_{AH} + r_F$	$rF + cFG + 2rBEG + cAB + rAH$
14 → 46	F-AH-BEG	GH	F-ABEGH4	AH + BEG	$D_{AH} + D_{BEG}$	$r_{AH} + r_{BEG}$	$rAH + cAB + rBEG$
15 → 16	F-H-ABEG	AH	F-ABEGH4	ABEG + H	$D_{ABEG} + D_H$	$r_{ABEG} + r_H$	$rABEG + cGH + rH$
15 → 31	F-H-ABEG	FH	FH-ABEG	F + H	$D_F + D_H$	$r_F + r_H$	$rF + cFG + 2rABEG + cGH + rH$
15 → 47	F-H-ABEG	GH	F-ABEGH4	ABEG + H	$D_{ABEG} + D_H$	$r_{ABEG} + r_H$	$rABEG + cGH + rH$
20 → 24	B-G-AEFH	BE	G-ABEFH	AEFH + B	$D_{AEFH} + D_B$	$r_{AEFH} + r_B$	$rB + cAB + rAEFH$
20 → 28	B-G-AEFH	BG	BG-AEFH	B + G	$D_B + D_G$	$r_B + r_G$	$rB + cAB + 2rAEFH + cFG + rG$
20 → 52	B-G-AEFH	GH	B-AEFGH	AEFH + G	$D_{AEFH} + D_G$	$r_{AEFH} + r_G$	$rG + cFG + rAEFH$
22 → 24	G-BE-AFH	AE	G-ABEFH	AFH + BE	$D_{AFH} + D_{BE}$	$r_{AFH} + r_{BE}$	$rAFH + cAB + rBE$
22 → 30	G-BE-AFH	BG	AFH-BGE	BE + G	$D_{BE} + D_G$	$r_{BE} + r_G$	$rBE + cAB + 2rAFH + cFG + rG$

(continued)

Table 1. Continued

Transition	Initial state MDs	Bond formed	Final state MDs	MDs coalescing	D	$R_{\min}$	$R_{\max}$
22 → 54	G-BE-AFH	GH	BE-AFGH	AFH + G	$D_{AFH} + D_G$	$r_{AFH} + r_G$	$rG + cFG + rAFH$
23 → 24	G-ABE-FH	AH	G-ABEFH	ABE + FH	$D_{ABE} + D_{FH}$	$r_{ABE} + r_{FH}$	$rABE + cEF + rFH$
23 → 31	G-ABE-FH	BG	FH-ABEG	ABE + G	$D_{ABE} + D_G$	$r_{ABE} + r_G$	$rABE + cEF + 2rFH + cFG + rG$
23 → 55	G-ABE-FH	GH	ABE-FGH	FH + G	$D_{FH} + D_G$	$r_{FH} + r_G$	$rG + cFG + rFH$
26 → 28	E-BG-AFH	AE	BG-AEFH	AFH + E	$D_{AFH} + D_E$	$r_{AFH} + r_E$	$rAFH + cEF + rE$
26 → 30	E-BG-AFH	BE	AFH-BGE	BG + E	$D_{BG} + D_E$	$r_{BG} + r_E$	$rBG + cAB + 2rAFH + cEF + rE$
26 → 58	E-BG-AFH	GH	E-ABFGH	AFH + BG	$D_{AFH} + D_{BG}$	$r_{AFH} + r_{BG}$	$rBG + cAB + rAFH$
27 → 28	AE-BG-FH	AH	BG-AEFH	AE + FH	$D_{AE} + D_{FH}$	$r_{AE} + r_{FH}$	$rAE + cEF + rFH$
27 → 31	AE-BG-FH	BE	FH-ABEG	AE + BG	$D_{AE} + D_{BG}$	$r_{AE} + r_{BG}$	$rBG + cAB + rAE$
27 → 59	AE-BG-FH	GH	AE-BFGH	BG + FH	$D_{BG} + D_{FH}$	$r_{BG} + r_{FH}$	$rBG + cFG + rFH$
29 → 30	A-FH-BEG	AH	AFH-BEG	A + FH	$D_A + D_{FH}$	$r_A + r_{FH}$	$rA + cAB + 2rBEG + cEF + rFH$
29 → 31	A-FH-BEG	AE	FH-ABEG	A + BEG	$D_A + D_{BEG}$	$r_A + r_{BEG}$	$rA + cAB + rBEG$
29 → 61	A-FH-BEG	GH	A-BEFGH	BEG + FH	$D_{BEG} + D_{FH}$	$r_{BEG} + r_{FH}$	$rBEG + cEF + rFH$
36 → 40	B-F-AEGH	BE	F-ABEGH4	AEGH + B	$D_{AEGH} + D_B$	$r_{AEGH} + r_B$	$rB + cAB + rAEGH$
36 → 44	B-F-AEGH	BG	F-ABEGH4	AEGH + B	$D_{AEGH} + D_B$	$r_{AEGH} + r_B$	$rB + cAB + rAEGH$
36 → 52	B-F-AEGH	FH	B-AEFGH	AEGH + F	$D_{AEGH} + D_F$	$r_{AEGH} + r_F$	$rF + cFG + rAEGH$
38 → 40	F-BE-AGH	AE	F-ABEGH4	AGH + BE	$D_{AGH} + D_{BE}$	$r_{AGH} + r_{BE}$	$rAGH + cAB + rBE$
38 → 46	F-BE-AGH	BG	F-ABEGH4	AGH + BE	$D_{AGH} + D_{BE}$	$r_{AGH} + r_{BE}$	$rBE + cAB + rAGH$
38 → 54	F-BE-AGH	FH	BE-AFGH	AGH + F	$D_{AGH} + D_F$	$r_{AGH} + r_F$	$rF + cFG + rAGH$
39 → 40	F-GH-ABE	AH	F-ABEGH4	ABE + GH	$D_{ABE} + D_{GH}$	$r_{ABE} + r_{GH}$	$rABE + cEF + 2rF + cFg + rGH$
39 → 47	F-GH-ABE	BG	F-ABEGH4	ABE + GH	$D_{ABE} + D_{GH}$	$r_{ABE} + r_{GH}$	$rABE + cEF + 2rF + cFg + rGH$
39 → 55	F-GH-ABE	FH	ABE-FGH	F + GH	$D_F + D_{GH}$	$r_F + r_{GH}$	$rF + cFG + rGH$
42 → 44	E-F-ABGH	AE	F-ABEGH4	ABGH + E	$D_{ABGH} + D_E$	$r_{ABGH} + r_E$	$rABGH + cFG + 2rF + cEF + rE$
42 → 46	E-F-ABGH	BE	F-ABEGH4	ABGH + E	$D_{ABGH} + D_E$	$r_{ABGH} + r_E$	$rABGH + cFG + 2rF + cEF + rE$
42 → 58	E-F-ABGH	FH	E-ABFGH	ABGH + F	$D_{ABGH} + D_F$	$r_{ABGH} + r_F$	$rF + cFG + rABGH$
43 → 44	F-AE-BGH	AH	F-ABEGH4	AE + BGH	$D_{AE} + D_{BGH}$	$r_{AE} + r_{BGH}$	$rAE + cAB + rBGH$
43 → 47	F-AE-BGH	BE	F-ABEGH4	AE + BGH	$D_{AE} + D_{BGH}$	$r_{AE} + r_{BGH}$	$rBGH + cAB + rAE$
43 → 59	F-AE-BGH	FH	AE-BFGH	BGH + F	$D_{BGH} + D_F$	$r_{BGH} + r_F$	$rF + cFG + rBGH$
45 → 46	A-F-BEGH	AH	F-ABEGH4	A + BEGH	$D_A + D_{BEGH}$	$r_A + r_{BEGH}$	$rA + cAB + rBEGH$
45 → 47	A-F-BEGH	AE	F-ABEGH4	A + BEGH	$D_A + D_{BEGH}$	$r_A + r_{BEGH}$	$rA + cAB + rBEGH$
45 → 61	A-F-BEGH	FH	A-BEFGH	BEGH + F	$D_{BEGH} + D_F$	$r_{BEGH} + r_F$	$rF + cFG + rBEGH$
50 → 52	B-E-AFGH	AE	B-ABFGH	AFGH + E	$D_{AFGH} + D_E$	$r_{AFGH} + r_E$	$rAFGH + cEF + rE$
50 → 54	B-E-AFGH	BE	BE-AFGH	B + E	$D_B + D_E$	$r_B + r_E$	$rB + cAB + 2rAFGH + cEF + rE$
50 → 58	B-E-AFGH	BG	E-ABFGH	AFGH + B	$D_{AFGH} + D_E$	$r_{AFGH} + r_E$	$rB + cAB + rAFGH$
51 → 52	B-AE-FGH	AH	B-AEFGH	AE + FGH	$D_{AE} + D_{FGH}$	$r_{AE} + r_{FGH}$	$rAE + cEF + rFGH$
51 → 55	B-AE-FGH	BE	ABE-FGH	AE + B	$D_{AE} + D_B$	$r_{AE} + r_B$	$rB + cAB + rAE$
51 → 59	B-AE-FGH	BG	AE-BFGH	B + FGH	$D_B + D_{FGH}$	$r_B + r_{FGH}$	$rB + cAB + 2rAE + cEF + rFGH$
53 → 54	A-BE-FGH	AH	BE-AFGH	A + FGH	$D_A + D_{FGH}$	$r_A + r_{FGH}$	$rA + cAB + 2rBE + cEF + rFGH$
53 → 55	A-BE-FGH	AE	ABE-FGH	A + BE	$D_A + D_{BE}$	$r_A + r_{BE}$	$rA + cAB + rBE$
53 → 61	A-BE-FGH	BG	A-BEFGH	BE + FGH	$D_{BE} + D_{FGH}$	$r_{BE} + r_{FGH}$	$rBE + cEF + rFGH$
57 → 58	A-E-BFGH	AH	E-ABFGH	A + BFGH	$D_A + D_{BFGH}$	$r_A + r_{BFGH}$	$rA + cAB + rBFGH$
57 → 59	A-E-BFGH	AE	AE-BFGH	A + E	$D_A + D_E$	$r_A + r_E$	$rA + cAB + 2rBFGH + cEF + rE$
57 → 61	A-E-BFGH	BE	A-BEFGH	BFGH + E	$D_{BFGH} + D_E$	$r_{BFGH} + r_E$	$rBFGH + cEF + rE$
16 → 32	F-ABEGH4	FH	ABEFGH5	ABEGH4 + F	$D_{ABEGH} + D_F$	$r_{ABEGH} + r_F$	$rF + cEF + rABEGH$
16 → 48	F-ABEGH4	GH	F-ABEGH5	(G + H) _{int}	$D_G + D_H$	$r_G + r_H$	prABEGH
24 → 32	G-ABEFH	BG	ABEFGH5	ABEFH + G	$D_{ABEFH} + D_G$	$r_{ABEFH} + r_G$	$rABEFH + cFG + rG$
24 → 56	G-ABEFH	GH	ABEFGH5	ABEFH + G	$D_{ABEFH} + D_G$	$r_{ABEFH} + r_G$	$rG + cFG + rABEFH$
28 → 32	BG-AEFH	BE	ABEFGH5	AEFH + BG	$D_{AEFH} + D_{BG}$	$r_{AEFH} + r_{BG}$	$rBG + cAB + rAEFH$
28 → 60	BG-AEFH	GH	ABEFGH5	AEFH + BG	$D_{AEFH} + D_{BG}$	$r_{AEFH} + r_{BG}$	$rBG + cAB + rAEFH$
30 → 32	AFH-BEG	AE	ABEFGH5	AFH + BEG	$D_{AFH} + D_{BEG}$	$r_{AFH} + r_{BEG}$	$rAFH + cAB + rBEG$
30 → 62	AFH-BEG	GH	ABEFGH5	AFH + BEG	$D_{AFH} + D_{BEG}$	$r_{AFH} + r_{BEG}$	$rAFH + cAB + rBEG$
31 → 32	FH-ABEG	AH	ABEFGH5	ABEG + FH	$D_{ABEG} + D_{FH}$	$r_{ABEG} + r_{FH}$	$rABEG + cEF + rFH$
31 → 63	FH-ABEG	GH	ABEFGH5	ABEG + FH	$D_{ABEG} + D_{FH}$	$r_{ABEG} + r_{FH}$	$rABEG + cEF + rFH$
40 → 48	F-ABEGH4	BG	F-ABEGH5	(B + G) _{int}	$D_B + D_G$	$r_B + r_G$	prABEGH
40 → 56	F-ABEGH4	FH	ABEFGH5	ABEGH4 + F	$D_{ABEGH} + D_F$	$r_{ABEGH} + r_F$	$rF + cEF + rABEGH$
44 → 48	F-ABEGH4	BE	F-ABEGH5	(B + E) _{int}	$D_B + D_E$	$r_B + r_E$	prABEGH
44 → 60	F-ABEGH4	FH	ABEFGH5	ABEGH4 + F	$D_{ABEGH} + D_F$	$r_{ABEGH} + r_F$	$rF + cEF + rABEGH$
46 → 48	F-ABEGH4	AE	F-ABEGH5	(A + E) _{int}	$D_A + D_E$	$r_A + r_E$	prABEGH
46 → 62	F-ABEGH4	FH	ABEFGH5	ABEGH4 + F	$D_{ABEGH} + D_F$	$r_{ABEGH} + r_F$	$rF + cEF + rABEGH$
47 → 48	F-ABEGH4	AH	F-ABEGH5	(A + H) _{int}	$D_A + D_H$	$r_A + r_H$	prABEGH
47 → 63	F-ABEGH4	FH	ABEFGH5	ABEGH + F	$D_{ABEGH} + D_F$	$r_{ABEGH} + r_F$	$rF + cEF + rABEGH$

(continued)

Table 1. Continued

Transition	Initial state MDs	Bond formed	Final state MDs	MDs coalescing	D	$R_{\min}$	$R_{\max}$
52 → 56	B-AEFGH	BE	ABEFGH5	AEFGH + B	$D_{AEFGH} + D_B$	$r_{AEFGH} + r_B$	$rB + cAB + rAEFGH$
52 → 60	B-AEFGH	BG	ABEFGH5	AEFGH + B	$D_{AEFGH} + D_B$	$r_{AEFGH} + r_B$	$rB + cAB + rAEFGH$
54 → 56	BE-AFGH	AE	ABEFGH5	AFGH + BE	$D_{AFGH} + D_{BE}$	$r_{AFGH} + r_{BE}$	$rAFGH + cAB + rBE$
54 → 62	BE-AFGH	BG	ABEFGH5	AFGH + BE	$D_{AFGH} + D_{BE}$	$r_{AFGH} + r_{BE}$	$rAFGH + cAB + rBE$
55 → 56	ABE-FGH	AH	ABEFGH5	ABE + FGH	$D_{ABE} + D_{FGH}$	$r_{ABE} + r_{FGH}$	$rABE + cEF + rFGH$
55 → 63	ABE-FGH	BG	ABEFGH5	ABE + FGH	$D_{ABE} + D_{FGH}$	$r_{ABE} + r_{FGH}$	$rABE + cEF + rFGH$
58 → 60	E-ABFGH	AE	ABEFGH5	ABFGH + E	$D_{ABFGH} + D_E$	$r_{ABFGH} + r_E$	$rABFGH + cEF + rE$
58 → 62	E-ABFGH	BE	ABEFGH5	ABFGH + E	$D_{ABFGH} + D_E$	$r_{ABFGH} + r_E$	$rABFGH + cEF + rE$
59 → 60	AE-BFGH	AH	ABEFGH5	AE + BFGH	$D_{AE} + D_{BFGH}$	$r_{AE} + r_{BFGH}$	$rAE + cAB + rBFGH$
59 → 63	AE-BFGH	BE	ABEFGH5	AE + BFGH	$D_{AE} + D_{BFGH}$	$r_{AE} + r_{BFGH}$	$rAE + cAB + rBFGH$
61 → 62	A-BEFGH	AH	ABEFGH5	A + BEFGH	$D_A + D_{BEFGH}$	$r_A + r_{BEFGH}$	$rA + cAB + rBEFGH$
61 → 63	A-BEFGH	AE	ABEFGH5	A + BEFGH	$D_A + D_{BEFGH}$	$r_A + r_{BEFGH}$	$rA + cAB + rBEFGH$
32 → 64	ABEFGH5	GH	ABEFGH6	$(G + H)_{\text{int}}$	$D_G + D_H$	$r_G + r_H$	prABEFGH
48 → 64	F-ABEGH5	FH	ABEFGH6	$ABEGH5 + F$	$D_{ABEGH} + D_F$	$r_{ABEGH} + r_F$	$rF + cEF + rABEGH$
56 → 64	ABEFGH5	BG	ABEFGH6	$(B + G)_{\text{int}}$	$D_B + D_G$	$r_B + r_G$	prABEFGH
60 → 64	ABEFGH5	BE	ABEFGH6	$(B + E)_{\text{int}}$	$D_B + D_E$	$r_B + r_E$	prABEFGH
62 → 64	ABEFGH5	AE	ABEFGH6	$(A + E)_{\text{int}}$	$D_A + D_E$	$r_A + r_E$	prABEFGH
63 → 64	ABEFGH5	AH	ABEFGH6	$(A + H)_{\text{int}}$	$D_A + D_H$	$r_A + r_H$	prABEFGH

The quantities λ_1 is the rate for the switching of a non-associated microdomain pair from a state in which both microdomains are correctly *folded* and *oriented* for coalescence to any other state and λ_2 is the rate for the reverse process. We have $\beta = \lambda_2/(\lambda_1 + \lambda_2)$.

We write $\alpha = 1/\sqrt{D\tau_c}$ where $1/\tau_c = \lambda_1 + \lambda_2$ is the characteristic rate of helix-coil transitions.

For the backward, unfolding rate $1/\tau_b$, the accessible surface area A_{AB} buried between microdomains A and B when they coalesce is used, as well as the free energy change per unit area f appropriate for this surface area measure (Lee & Richards, 1971). The parameter ν determines the degree to which the folding process goes to completion. The equation for τ_b is

$$\tau_b = \nu^{-1} e^{\frac{f A_{AB}}{k_B T}}. \quad (5)$$

The chemical kinetics approximation to diffusion-collision folding kinetics yields a set of coupled first-order rate equations for the probabilities of each microdomain or multimicrodomain complex (corresponding to concentrations in unimolecular processes) $p_i(t)$ of the possible intermediate states which are

$$\frac{dp_i}{dt} = \sum_{j=1}^{64} K_{ij} p_j. \quad (6)$$

In Equation 3, K_{ij} are the elements of the rate matrix to be determined from two-microdomain model calculations. They are determined by the individual two-microdomain rates described above in Equations 1 and 2 (see Karplus & Weaver, 1994; Yapa & Weaver, 1996 for details).

Simulation 1: Consistent with recent experiments

The results of Simulation 1 of apomyoglobin folding kinetics are shown in Figure 3A–C. The probabilities (corresponding to concentrations for this unimolecular process) of the prominent kinetic intermediates are plotted versus time. Besides the initial (state 1) and final (state 64) states whose time development is shown in Figure 3A, the only intermediates which accumulate significantly are state 33 (GH), state 34 (AGH), and state 42 (ABGH) shown in

Fig. 2. A: Schematic diagram of the interaction of helical microdomains G and H, showing the minimum and maximum distances of separation. B: Example of the accessible surface areas of two multimicrodomain complexes relevant to the G-H interaction.

Fig. 3. The probabilities as a function of time for the kinetic intermediates of apomyoglobin when the properties of the individual helical microdomains dominate the folding kinetics.

Figure 3B. Their prominence in the folding pathways comes from choosing the probability parameter β for each of the 192 transitions in accordance with the recent experimental results on apomyoglobin folding in which the AGH and ABGH complexes appear early in the folding, with AGH appearing in about $7 \mu\text{s}$ (Ballew et al., 1996). The parameter β is the probability that a microdomain (or multimicrodomain complex) will favorably interact with another microdomain when their mutual diffusion leads to their collision.

If folding were to proceed only by diffusion-controlled collisions between microdomains (i.e., if only the first term in Equation 1 was important), then the A and H microdomains would coalesce in about $2 \mu\text{s}$. This is the diffusion-collision limit for these helices at the N and C-terminal ends of the apomyoglobin chain in the chemical kinetics approximation. Under the same circumstances, the G and H microdomains would coalesce in about 4 ns (diffusion-collision limit). However, once the GH pair has formed, the AGH cluster can coalesce in about $1 \mu\text{s}$ because the AH distance has been reduced. Since the G and H helices are adjacent on the chain, their early coalescence is favored by geometrical considerations. However, the time for their coalescence is also influenced by the second term in Equation 1, which contains

the coalescence probability parameter β . In the part of Simulation 1 shown in Figure 3B, the choice of β for the transition to the GH kinetic intermediate (state 33) leads to a GH pair formation time of about $2 \mu\text{s}$. Once the GH pair has formed, it is probable that the H helix residues which interact with A will be more likely to be in a helical conformation, making the AH coalescence more probable. Note in Figure 3B that the three kinetic intermediates GH, AGH, and ABGH occupy, in turn, a substantial portion of the available probability for the molecule (more than 40% in the case of AGH), so they should be detectable in fast kinetics experiments (Ballew et al., 1996).

In the apomyoglobin simulations, because of the complex relationship between the kinetic intermediates described in Fig. 1, the AE helix pairing must be suppressed in addition to the AH pairing being enhanced in order to have AGH appear as a prominent kinetic intermediate. Suppressing the formation of the AE kinetic intermediate with a relatively small β for E also suppresses the formation of the BE pair relative to the BG pairing in ABGH (state 42). This leads to state 42 being a prominent kinetic intermediate, peaking around $17 \mu\text{s}$ as shown in Figure 3B.

The FH pairing must also be suppressed in apomyoglobin folding by reducing the stability of the F helix relative to the A helix.

Otherwise, due to the favorable diffusional time scale for near neighbor collisions, the FGH complex (state 49) will appear prominently. In fact, a likely alternative pathway (instead of through AGH) to the folded state 64 would probably proceed through the FGH complex (state 49) to the AFGH complex (state 50) and then to the ABFGH complex (state 58) shown in Figure 3C.

In this simulation, we have shown that through appropriate choices of forward and reverse reaction rate constants for pairwise association and dissociation, early intermediates along the folding pathway of apomyoglobin can arise out of the association of the A, G, and H helices. The adjustable parameters are the folding probabilities β_f of the individual microdomains and the orientational probabilities β_a determined by the packing of the helices. The orientational probability is related to the hydrophobic effect or the free energy change upon solvation for the individual microdomains (Weaver, 1992). The folding properties β_f are determined by the helix-coil equilibria of the individual helical microdomains. To investigate the helical stabilities in apomyoglobin, Hirst and Brooks (1995) performed molecular dynamics studies on the six helices involved in our chemical kinetics simulations. From analyses of hydrogen bonding and fluctuations from simulations at 298 and 368 K, they found the relative stabilities of the helices to be $A \approx G \approx H > B > E > F$, consistent with our requirements for β to produce the AGH and ABGH kinetic intermediates early and in high concentration in the folding of apomyoglobin. Since the kinetics depend on the properties of individual microdomains (helices in this case), we are beginning a further study of the myoglobin helices using the electrostatic multipole representation (Pappu et al., 1996) to extract values for β .

Simulation 2: Geometry-based folding kinetics

To contrast with the experimental results and with Simulation 1, we performed a simulation in which all of the microdomains of the same degree of complexity (elementary, one-pair, etc.) had identical β s and with the nascent helices A-to-H being only marginally stable. This simulation allows us to assess the effects of proximity of microdomains in the chain on the folding kinetics and is related to the earlier simulations of Bashford et al. (1988). The results of this simulation are shown in Figure 4. Here, we notice that, following formation of GH, if geometry (proximity) were the only criterion affecting folding kinetics, the FGH cluster would appear as the most prominent kinetic intermediate rather than the AGH cluster. Since the experimental evidence suggests that the AGH cluster is the most prominent kinetic intermediate, the properties of individual helical microdomains in apomyoglobin must suppress the association of the F helix with the GH pair and promote the association of the A helix with the GH pair. The most likely parameter to affect coalescence upon collision is β , which is related to helix-coil equilibrium constants in the case of helical microdomains and to orientational effects upon collision.

Much work remains to be done in order to establish equilibrium constants and the time scale for the transitions of helical and other microdomains. Further molecular dynamics simulations such as those of Hirst and Brooks (1995) are needed to improve our knowledge of microdomain properties. A possible approach is to incorporate transiently folding microdomains into Brownian dynamics (Ermak & McCammon, 1978; McCammon et al., 1980; Yapa et al., 1992; Schneller & Weaver, 1993; Yapa & Weaver, 1996) or other reduced (Pappu & Weaver, 1997) or simplified representation simulations, since the former are already into the microsecond

Fig. 4. The probabilities as a function of time for the kinetic intermediates of apomyoglobin when nearest neighbor interactions dominate the folding kinetics.

time regime (McCammon, 1996; Yapa & Weaver, 1996) in which important folding kinetic events take place. The recent work of Hoffmann and Knapp (1997) illustrates the utility of this approach.

Discussion

The diffusion-collision model of protein folding (Karplus & Weaver, 1976; Karplus & Weaver, 1994; Yapa & Weaver, 1996) has been applied to simulate the early folding kinetics of apomyoglobin by solving a set of first-order rate equations for the probabilities of the 64 folding states. The number of states is determined by the number of helix-helix contacts considered important in the kinetic pathways. In this case, the six most significant hydrophobic interactions between the A, B, E, F, G, and H helices (Weaver, 1992) have been taken as the helix-helix pairings whose kinetics of formation determine the early folding paths. The two alternatives for the kinetic pathway found in Simulations 1 and 2 are shown schematically in Figure 5. In the upper path, which is consistent with experiment, a molecule goes from the unfolded state through the GH pairing to the AGH complex with the GH and AH pairings made and subsequently through the ABGH complex with the addition of the BG pairing. This is the phenomenological limit to the DC folding of apomyoglobin. In the purely diffusion controlled limit, the most likely next pairing is FH followed by the AE and BE pairs leading to the final state. In folded apomyoglobin, there are additional smaller pairings among the six helices as well as contacts between B and the small C and D helices. If F were made more stable and A less stable initially by selective point mutations, this pathway would become relatively more important.

Since the folding of apomyoglobin and other globular proteins involves diffusion of segments of the polypeptide chain together, changing the distances between interacting segments (microdomains) will change the extent of diffusion control on folding pathways. For example, if the distance between helices E and F is made larger (say add ten residues in a random coil to the chain connecting them), the major pathway through the AGH cluster can

Fig. 5. Schematic diagram of the important folding pathways in apomyoglobin.

be enhanced in comparison to the FGH pathway. Similarly, if the distance between helices G and H is made larger by ten residues, the major pathway would be suppressed. In general, changing the distance between microdomains that fold together will probably alter the likelihood of formation of some kinetic intermediates without affecting the final fold. It may be possible by changing the distances to make some kinetic intermediates more accessible to experimental investigation.

The importance of spatial proximity of microdomains in determining the kinetic pathways in folding underscores an important consequence of early formation of S-S bonding (in addition to increased stability). Early S-S bonds will alter folding kinetics by decreasing diffusion distances between microdomains which are far apart on the linear chain. Conversely, lack of early S-S bonds in a protein implies that stabilities of microdomains which are far apart in sequence play a greater role in the folding kinetics (for example, apomyoglobin in which the results of our simulations indicate that helices A, G, and H must be substantially more stable in isolation than helices E and F and more stable than helix B).

In our diffusion-collision chemical kinetics study of apomyoglobin folding intermediates, we set all of the orientational folding probabilities β_a to one and did not investigate how changes in their values would affect the magnitude and time scale of the population of kinetic intermediates. In fact, it is probably true that in the early stages of folding, the hydrophobic interaction between microdomains is different from that found in the completely folded structure. Evidence for this possibility comes from comparing experimental amide proton protection rates for individual apomyoglobin residues (Shin et al., 1993) with the accessible area losses of those residues involved in the native AGH cluster (Weaver, 1992). Jennings and Wright (1993) found that residues A (L9, V10, L11, W14, V17), residues G (K102, L104, I107, A110, I112, H113, V114, L115), and residues H (K133, F138, I142, A143) were quickly protected. A number of the hydrophobic contacts found in the native structure (Weaver, 1992) are not made in the AGH intermediate. In addition, there are a few quickly protected residues (e.g., G(K102, I107)) that are not in contact in the native structure.

The results of the simulations summarized in Figure 3A–C provide a diffusion-collision model based explanation of the kinetic experimental results of Ballew et al. (1996). The model explains the viscosity dependence of the kinetics of folding of apomyoglobin through the diffusion coefficients D in the folding rates. An alternative way to alter the kinetics would be by changing the chain length between microdomain pairs important on the main folding pathways as discussed above.

The diffusion-collision model results are also consistent with the equilibrium experiments of Loh et al. (1995) who observed a sec-

ond intermediate with more structure than the AGH intermediate found previously (Hughson et al., 1990, 1991; Jennings & Wright, 1993; Jennings et al., 1994). The new intermediate has the B helix in addition to A, G, and H and corresponds to state 42 in our simulations, namely, ABGH with the AH, BG, and GH pairings made. Again, in the intermediate of Loh et al. (1995), not all the hydrophobic contacts (as determined by accessible area loss) seen in the native structure are present in this intermediate (as determined by protection factors).

If all betas are set equal to one, that is, all helices are completely folded and correctly oriented at each collision (results not shown and Bashford et al., 1988), then the GH pair and FGH triplet form quickly with a time scale of the order of 10 ns. This time scale is consistent with the results of Brownian dynamics simulations of the movements and collisions of helix-turn-helix (Lee et al., 1987) and four-helix bundle (Yapa & Weaver, 1996) motifs. It is smaller than the estimates of Hagen et al. (1996) for the lower limit for the time of diffusion-collision of residues in cytochrome *c* under strongly denaturing conditions.

The methods described in this paper illustrate the first two stages in an algorithm for solving the folded protein structure prediction problem. The first stage is the model-building of a set of quasi-particles from the sequence of the polypeptide. This stage produces a possible set of microdomains and allows, in principle, the calculation of their properties. In this work, the set was taken to be the nascent forms of the six important helices seen in the native structure of apomyoglobin. The second stage is the folding dynamics which brings the protein to the loosely associated state at the end of the kinetic folding pathway (molten globule). In this paper, the kinetics are supplied by the chemical kinetics approximation to the diffusion-collision model. In the third stage, which is not considered in this work, the molten globule kinetic final state (state 64) will attain the close packing observed in the native structure. A computational combination of energy minimization and molecular dynamics on a full representation of the polypeptide chain is required for this stage.

Materials and methods

As computed in Table 1, $R_{\max}$ is the fully extended length of the connecting chain between two microdomains taken as 3.5 Å times the number of intervening coil residues plus the distances on each end to the two microdomain centers. If there are helical microdomains along the connecting chain, their contribution to the length is taken to be the diameter of the microdomain in the spherical approximation. For transitions involving multimicrodomain clusters, the distance $R_{\max}$ is taken as the shortest chain between the interacting microdomains.

In apomyoglobin folding, there are also *internal* transitions involving the association of microdomains that are in the same multimicrodomain complex. The necessity of internal transitions to complete the folding of a globular protein gives a physical reason for molten globule behavior in folding proteins. The kinetic folding intermediates must retain flexibility in order for these transition to take place. It also means that $R_{\max}$ must be modified from the example above to apply to this class of folding transitions. The time scale for the internal transitions will depend on the geometry of the multimicrodomain complex. For the folding time of internal transitions, we also use Equation 1 where $R_{\min}$ is as defined for external transitions and $R_{\max}$ is now half the circumference of the multimicrodomain complex in the spherical approximation. D is the relative diffusion coefficient for the pair of microdomains involved in the internal transition. This gives an upper limit to the folding times in these cases. The internal transitions will readily take place in a molten globule-like intermediate state in which the prior microdomain-microdomain associations are still loose enough for movements to occur. Full native-like interdigitation of the sidechains involved in hydrophobic interactions between the helices would fix the sidechains in a particular orientation by pair-wise interactions and preclude internal transitions.

The parameter β in Equation 1 determines the probability that in a collision between microdomains, coalescence occurs. The β for each of the 192 transitions is a product of four individual microdomain probabilities, two for each microdomain or multimicrodomain involved in the pairwise interaction. Each member of an interacting pair has an orientational probability and a folding probability. The orientation probabilities are parameters β_{a_i} and β_{a_j} for microdomains i and j which are defined as the ratio of the lost accessible area to the total accessible area of the elementary microdomain or microdomain cluster. For example, in the transition $10 \rightarrow 42$, which involves the coalescence $AH + BG$ and the interaction of the pair GH (see Fig. 2B), $\beta_{a_i} = \Delta H_{GH}/A_{AH} = 0.109$ and $\beta_{a_j} = \Delta G_{GH}/A_{BG} = 0.135$.

The folding probabilities β_{f_i} and β_{f_j} are the probabilities that the microdomains are correctly folded when they collide. For example, for helical microdomains, these probabilities could be helix-coil equilibrium constants which are generally small for isolated helical fragments in aqueous solution (Dyson et al., 1992a, 1992b; Reymond et al., 1997).

The resulting β is a number between zero and one given by

$$\beta = \beta_{a_i} \beta_{a_j} \beta_{f_i} \beta_{f_j} \quad (7)$$

Since the orientational probabilities β_{a_i} and β_{a_j} tend to be relatively small, β will be smaller than one in all transition (except internal transitions). An alternative possibility that we used previously (Bashford et al., 1988; Yapa & Weaver, 1996) is that all orientation probabilities are closer to one than found from the above analysis, because surface diffusion or reorienting of microdomains occurs rapidly after collision and loose coalescence by correctly folded microdomains. This would be suggested by molten globule-like behavior in kinetic intermediates. Further study of surface diffusion on protein surfaces is required to clarify the time scale of surface rearrangements of microdomains in multi-microdomain clusters. The recent helix-turn-helix molecular dynamics simulations of Hoffmann and Knapp (1997) displayed the surface diffusion phenomenon with the two helices.

To complete the calculation of β for each forward transition, we must choose values for the folding probabilities β_{f_i} and β_{f_j} in

Equation 4. In the case of apomyoglobin, the folding β s will be related to the probabilities of the A, B, E, F, G, and H amino acid groupings being in conformations that lead to coalescence upon collision. Clearly, the choices for β will have a substantial effect on the probabilities of occurrence of each of the kinetic states as a function of time.

The kinetic calculations, mainly solving Equation 3, have been carried out with MATLAB (1996), a numerical analysis software product, well suited to the matrix calculations required, and having the advantage of running on most computer hardware platforms. The MATLAB code for the apomyoglobin calculations is available from the authors.

Acknowledgment

The authors would like to thank Martin Karplus for useful comments.

References

- Baldwin RL. 1996. Why is protein folding so fast? *Proc Natl Acad Sci USA* 93:2627–2628.
- Ballew RM, Sabelko J, Gruebele M. 1996. Direct observation of fast protein folding: The initial collapse of apomyoglobin. *Proc Natl Acad Sci USA* 93:5759–5764.
- Bashford D, Cohen FE, Karplus M, Kuntz ID, Weaver DL. 1988. Diffusion-collision model for the folding kinetics of myoglobin. *Proteins* 4:211–227.
- Bernstein FC, Koetzle TF, Williams GJB, Meyer EF, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. 1977. The protein data bank: A computer based archival file for macromolecular structure. *J Mol Biol* 112:535–542.
- Brooks CL, Karplus M, Pettitt BM. 1988. *Proteins: A theoretical perspective of dynamics, structure and thermodynamics*. New York: Wiley.
- Dyson HJ, Merutka G, Waltho JP, Lerner RA, Wright PE. 1992. Folding of peptide fragments comprising the complete sequence of proteins. Models for initiation of protein folding I. Myohemerythrin. *J Mol Bio* 226:795–817.
- Dyson HJ, Sayre JR, Merutka G, Shin H, Lerner RA, Wright PE. 1992. Folding of peptide fragments comprising the complete sequence of proteins. Models for initiation of protein folding II. Plastocyanin. *J Mol Bio* 226:819–835.
- Eliczer D, Jennings PA, Wright PE, Doniach S, Hodgson KO, Tsuruta H. 1995. The radius of gyration of an apomyoglobin folding intermediate. *Science* 270:487–488.
- Ermak DL, McCammon JA. 1978. Brownian dynamics with hydrodynamic interactions. *J Chem Phys* 69:1352–1360.
- Hagen SJ, Hofrichter J, Szabo A, Eaton WA. 1996. Diffusion-limited contact formation in unfolded cytochrome c: Estimating the maximum rate of protein folding. *Proc Natl Acad Sci USA* 93:11615–11617.
- Hirst JD, Brooks CL III. 1995. Molecular dynamics simulations of isolated helices of myoglobin. *Biochemistry* 34:7614–7621.
- Hoffmann D, Knapp E-W. 1997. Folding pathways of a helix-turn-helix model protein. *J Phys Chem*. Forthcoming.
- Hughson FM, Wright PE, Baldwin RL. 1990. Structural characterization of a partly folded apomyoglobin intermediate. *Science* 249:1544–1548.
- Hughson FM, Barrick D, Baldwin RL. 1991. Probing the stability of a partly folded apomyoglobin intermediate by site-directed mutagenesis. *Biochemistry* 30:4113–4118.
- Jacob M, Schindler T, Balbach J, Schmid. 1997. Diffusion control in an elementary protein folding reaction. *Proc Natl Acad Sci USA* 94:5622–5627.
- Jennings PA, Wright PE. 1993. Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. *Science* 262:892–896.
- Jennings PA, Dyson HJ, Wright PE. 1994. In: S. Doniach, ed. *The folding pathway of apomyoglobin, in statistical mechanics, protein structure and protein substrate interactions*. New York: Plenum Press. pp 7–18.
- Karplus M, Weaver DL. 1976. Protein-folding dynamics. *Nature* 260:404–406.
- Karplus M, Weaver DL. 1994. Protein-folding dynamics: The diffusion-collision model and experimental data. *Protein Sci* 3:650–669.
- Lee B, Richards FM. 1971. The interpretation of protein structure: Estimation of static accessibility. *J Mol Bio* 55:379–400.
- Lee SY, Karplus M, Bashford D, Weaver DL. 1987. Brownian dynamics simulation of protein folding: A study of the diffusion-collision model. *Biopolymers* 26:481–506.
- Levinthal C. 1968. Are there pathways for protein folding? *J Chim Phys* 65: 44–45.

- Levinthal C. 1966. Molecular model building by computer. *Scient Am* 214: 42–45.
- Loh SN, Kay MS, Baldwin RL. 1995. Structure and stability of a second molten globule intermediate in the apomyoglobin folding pathway. *Proc Natl Acad Sci USA* 92:5446–5450.
- MATLAB. 1996. *The language of technical computing*. Natick, MA: The Math-Works, Inc.
- McCammon JA, Northrup SH, Karplus M, Levy RM. 1980. Helix-coil transitions in a simple polypeptide model. *Biopolymers* 19:2033–2045.
- McCammon JA. 1996. A speed limit for protein folding. *Proc Natl Acad Sci USA* 93:11426–11427.
- Pappu RV, Schneller W, Weaver DL. 1996. Electrostatic multipole representation of a polypeptide chain: An algorithm for simulation of polypeptide properties. *J Comp Chem* 17:1033–1055.
- Reymond MT, Merutka G, Dyson HJ, Wright PE. 1997. Folding propensities of peptide fragments of myoglobin. *Protein Sci* 6:706–716.
- Schneller W, Weaver D. 1993. Simulation of α -helix-coil transitions in simplified polyvaline: Equilibrium properties and Brownian dynamics. *Biopolymers* 33:1519–1535.
- Shin H-C, Merutka G, Waltho JP, Wright PE, Dyson HJ. 1993. Peptide models of protein folding initiation sites. 2. The G-H turn region of myoglobin acts as a helix stop signal. *J Am Chem Soc* 115:6348–6355.
- Shin H-C, Merutka G, Waltho JP, Tennant LL, Dyson HJ, Wright PE. 1993. Peptide models of protein folding initiation sites. 3. The G-H helical hairpin of myoglobin. *J Am Chem Soc* 115:6356–6364.
- Waldburger CD, Jonsson T, Sauer RT. 1996. Barriers to protein folding: Formation of buried polar interactions is a slow step in acquisition of structure. *Proc Natl Acad Sci USA* 93:2629–2643.
- Waltho JP, Feher VA, Merutka G, Dyson HJ, Wright PE. 1993. Peptide models of protein folding initiation sites. 1. Secondary structure formation by peptides corresponding to the G- and H-helices of myoglobin. *J Am Chem Soc* 115:6337–6347.
- Weaver DL. 1992. Hydrophobic interaction between globin helices. *Biopolymers* 32:477–490.
- Yapa K, Weaver DL, Karplus M. 1992. β -Sheet-coil transitions in a simple polypeptide model. *Proteins* 12:237–265.
- Yapa KK, Weaver DL. 1996. Protein folding dynamics: Application of the diffusion-collision model to the folding of a four-helix bundle. *J Phys Chem* 100:2498–2509.