Brownian dynamics simulation of protein folding: A study og the diffusion colliosn model

ARTICLE *in* BIOPOLYMERS · APRIL 1987

Impact Factor: 2.39 · DOI: 10.1002/bip.360260404 · Source: PubMed

4 AUTHORS, INCLUDING:

Donald Bashford
St. Jude Children's Research Hospital
85 PUBLICATIONS **9,988** CITATIONS

SEE PROFILE

35981280

Request # 35981280 **APR 07, 2014**

Ariel To: ariel.stjude.org

St Jude Children's Research Hospital
Biomedical Library
262 Danny Thomas Place
Memphis, TN 38105-3678

DOCLINE: Journal Copy

Title:	Biopolymers
Title Abbrev:	Biopolymers
Citation:	1987 Apr;26(4):481-506
Article:	Brownian dynamics simulation of protein folding: a study of
Author:	Lee SY;Karplus M;Bashford D;Weaver D
NLM Unique ID:	0372525 Verify: PubMed
PubMed UI:	3567323
ISSN:	0006-3525 (Print) 1097-0282 (Electronic)
Fill from:	**Any format**
Publisher:	Wiley Interscience, New York, NY :
Copyright:	Copyright Compliance Law
Authorization:	Vanessa Hays
Need By:	N/A
Maximum Cost:	**$25.00**
Patron Name:	Bashford, Donald 54120
Referral Reason:	Not owned (title)
Library Groups:	AMAHSL
Phone:	1.901.595-3389
Fax:	..
Email:	library@stjude.org
Ariel/Odyssey:	Ariel: ariel.stjude.org
Alt Delivery:	Ariel,Email(PDF),Email(TIFF),Web(PDF),Web(TIFF)
Routing Reason:	Received Apr 07, 2014 10:42 by PAUICR in Serial Routing - cell 1

Bill to: TNUTSJ

St Jude Children's Research Hospital
Biomedical Library
262 Danny Thomas Place
Memphis, TN 38105-3678

Brownian Dynamics Simulation of Protein Folding: A Study of the Diffusion-Collision Model*

SANGYOUB LEE,† MARTIN KARPLUS, *Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138;* DONALD BASHFORD,‡ and DAVID WEAVER, *Department of Physics, Tufts University, Medford Massachusetts 02155*

Synopsis

The dynamic aspects of protein folding are described by a series of diffusion-collision steps involving structural units (microdomains) of various sizes that combine to form the protein in its native state. A method is introduced for obtaining the rate constants for the basic diffusion-collision step by use of Brownian dynamics. The method is applied to an investigation of the folding dynamics of two α-helices connected by a flexible (random-coil) polypeptide chain. The results of this full three-dimensional treatment are compared with simplified model calculations for the diffusion-collision step. Of particular interest are the nature of the collision dynamics and the role of the intervening peptide chain.

INTRODUCTION

Most analyses of the dynamics of the folding of globular proteins have been concerned with possible mechanisms for reducing the folding time from that required for a random search through all possible structures to a process that can take place in the range of the experimental estimates (10^{-3}–10^{2} s). The diffusion-collision model[1,2] is one such proposed mechanism in which the elementary step of the dynamics is the mutual diffusion and collision of two (possibly unstable) quasi-particles called "microdomains." Collisions between the microdomains sometimes lead to their coalescence into larger microdomain aggregates and eventually to the native structure of the protein. The elementary microdomains in the unfolded protein are local structures short enough for all conformational alternatives to be searched through rapidly. Possible elementary microdomains are helices, segments of β-sheets, hydrophobic residue aggregates, or other local arrangements of amino acid residues of transient stability that may be rapidly formed and dissociated.

Because of the long time required for folding, compared to the present extent of protein dynamics simulations with a realistic representation of the protein (on the order of nanoseconds), it is not possible to perform a detailed

*Supported in part by a grant from the National Science Foundation.
†Present address: Department of Chemistry, University of Colorado, Boulder, Colorado.
‡Present address: Department of Chemistry, Harvard University, Cambridge, Mass., 02138.

Biopolymers, Vol. 26, 481–506 (1987)

CCC 0006-3525/87/040481-26$04.00

simulation of the folding process at the atomic level. An alternative is to reduce the folding problem to a network of quasi-two-body steps using the available stereochemical information about the packing of secondary structural units in the native protein.* Then the rate constants for each of the elementary steps in the network can be calculated by analytical and/or numerical methods. Finally, the overall folding kinetics can be approximated by solving numerically kinetic equations that couple the elementary steps.

The purpose of the present paper is to examine a possible elementary step in the early stages of the folding of a helical protein, namely the diffusion and collision of two helical moieties at the ends of a polypeptide chain. Previous work[1,2,4-10] on this problem has involved modeling the helices by spheres and the intervening chain segments as a featureless string. In this paper, we perform Brownian dynamics simulations in which each amino acid residue is approximated by a single interaction center linked by virtual bonds. This allows us to assess the validity of some of the approximations of the analytical models (e.g., the effect of intervening chain segments on the diffusion of the helices) as well as to obtain detailed kinetic information on the elementary folding step between two short helices.

In the methodology section, the model used to represent the helices and the intervening chain is described and the Brownian dynamics method used to simulate the folding is summarized. Also, the method used to extract the kinetic information from the Brownian dynamics trajectories of the model peptide chain is outlined. The section on the analytic model considers the analytic treatment of the folding dynamics and provides a formulation that can be compared with the present results. These are presented and analyzed in the results and discussion section.

METHODOLOGY

In this section we describe the model used to represent the polypeptide chain and outline the Brownian dynamics method used to calculate the folding process.

Polypeptide Model

The system studied in the simulations is a 24-residue polypeptide, in which the first and last eight residues are constrained to be α-helical and the intervening eight residues are in random-coil conformation. The model used is the one employed in the investigation of the helix–coil transition problem by McCammon et al.[11] Each amino acid residue is represented by a single interaction center. These centers are linked by virtual bonds with constrained bond lengths and virtual bond angles with harmonic restoring potentials. The virtual dihedral angle ϕ_i is defined in terms of the centers of residues $i-3$, $i-2$, $i-1$, and i ($\phi_i = 0$ for the eclipsed conformation), and is associated with an intrinsic torsional potential. Interactions between residues separated by three or more virtual bonds are given by central force potentials that represent excluded-volume and net attractive effects in aqueous surroundings.

* For the folding kinetics of the λ-repressor operator-binding domain, for example, see Ref. 3.

To stabilize the two α-helices, additional torsional potentials are introduced; they can be regarded as corresponding to the backbone hydrogen bonds. Finally, the possibility of a dipole–dipole interaction between the helices,[12,13] due to the C=O and N—H dipoles, is represented by assigning a point dipole to each helical residue. The sum of all these potential terms constitutes the effective energy function for the polypeptide chain.

For the initial studies described here, the geometric and energy parameters of the model are chosen to correspond to polyvaline; generalization to other homopolymers or to heteropolymers is straightforward. The energy function is chosen to approximate the potential of mean force, which corresponds to a thermal average over polypeptide degrees of freedom omitted in the simplified model and over solvent molecule degrees of freedom. The energy function is written as a sum of terms in the form

$$V = V_b + V_\theta + V_\phi + V_\alpha + V_{ev} + V_{sol} + V_{dd} \tag{1}$$

The term V_b is associated with virtual bonds. In the present simulations, the virtual bond lengths are constrained to be equal to 5.14 Å by applying the SHAKE algorithm[14] so that V_b is a constant. Constraining the bond lengths alters the long-time dynamical properties only very slightly, but makes possible an increase of the time-step size in Brownian dynamics simulation.[15,16] The bond-angle interactions are harmonic; i.e.,

$$V_\theta = \sum_{\text{all } \theta} k_\theta(\theta - \theta_0)^2 \tag{2}$$

where $k_\theta = 40$ kcal mol^{-1} rad^{-2} and $\theta_0 = 1.52$ rad $= 87.2°$. The dihedral angle term has the form

$$V_\phi = \sum_{\text{all } \phi} f(\phi) \tag{3}$$

where $f(\phi)$ is the Fourier series for Ala-Ala dipeptide given by Levitt.[17] The helix-stabilization energy, which is applied only to the first eight and final eight residues, is

$$V_\alpha = \sum_{\text{all } \phi} A(\phi)h(\phi) \tag{4}$$

where $A(\phi_i) = -6.0$ kcal/mol if ϕ_i belongs to helical segments and $A(\phi_i) = 0$ otherwise. The function $h(\phi)$ is

$$h(\phi) = \begin{cases} 55.73\left[0.7854(\phi - 0.9599)^2 + 2.0(\phi - 0.9599)^3\right] & \text{if } 40° < \phi < 55° \\ 55.73\left[0.7854(\phi - 0.4363)^2 - 2.0(\phi - 0.4363)^3\right] & \text{if } 25° < \phi < 40° \\ 0.0 & \text{otherwise} \end{cases} \tag{5}$$

where ϕ is in radians. The excluded-volume term is given by

$$V_{ev} = \sum_{i<j} v_{ev}(r_{ij}) \tag{6}$$

where the summation runs over all pairs of residues separated by three or more bonds, r_{ij} is the distance between residues i and j, and

$$v_{ev}(r) = \begin{cases} \varepsilon\left[3\left(\frac{r^0}{r}\right)^8 - 4\left(\frac{r^0}{r}\right)^6 + 1\right] & \text{if } r < r^0 \\ 0 & \text{if } r > r^0 \end{cases} \tag{7}$$

with $\varepsilon = 0.33$ kcal/mol and $r^0 = 6.5$ Å. This contribution is purely repulsive to avoid overcounting the attractive interaction.[11] The attractive van der Waals and solvent contribution has the ad hoc form used by Levitt for valine residues[17]

$$V_{sol} = \sum_{i<j} \sigma g(r_{ij}) \tag{8}$$

where the summation runs over all pairs of residues separated by three or more bonds, $\sigma = -3.0$ kcal/mol, and $g(r)$ is a sigmoid function[11,17] that varies from $g(0) = 1$ to $g(r) = 0$ for $r > 9$ Å. The dipole–dipole interaction energy between the helical parts is given by

$$V_{dd} = \sum_i \sum_j E(\mathbf{r}_{ij}, \mu_i, \mu_j) \tag{9}$$

where the first summation runs over all residues in one of the helices and the second summation over all residues in the other, and

$$E(\mathbf{r}_{ij}, \mu_i, \mu_j) = \frac{1}{4\pi\epsilon(\mathbf{r}_{ij})} \frac{1}{r_{ij}^3}\left[(\mu_i \cdot \mu_j) - 3(\mu_i \cdot \hat{r}_{ij})(\mu_j \cdot \hat{r}_{ij})\right] \tag{10}$$

with $r_{ij} = |\mathbf{r}_{ij}|$ and $\hat{r}_{ij} = \mathbf{r}_{ij}/r_{ij}$. We have neglected the dipole–dipole interactions from the dipoles imbedded in the residues of the intervening random-coil part, and have assumed that the dipole–dipole interactions between the residues in the same helical part are included in the helix-stabilization energy V_α. The direction of the electric dipole vector μ_i of a residue is parallel to the helical axis to which the residue belongs,[12,13] as expected from the orientation of the C=O and NH groups. The dipole moment of a residue is set equal to 3.5 D. In Eq. (10) we introduce a dielectric function $\epsilon(\mathbf{r}_{ij})$ that interpolates between the dielectric constants of water ($\epsilon_w = 81$) and of proteins ($\epsilon_p \approx 2.5$) (see Fig. 1):

$$\epsilon(\mathbf{r}_{ij}) \approx \epsilon(r_{ij}) \approx \begin{cases} \epsilon_p + (\epsilon_w - \epsilon_p)\dfrac{(r_{ij} - 6)^2}{16 + (r_{ij} - 6)^2} & \text{if } r_{ij} \geq 6\,\text{Å} \\ \epsilon_p & \text{if } r_{ij} < 6\,\text{Å} \end{cases} \tag{11}$$

Brownian Dynamics Algorithm

To investigate the folding dynamics of the 24-residue polypeptide chain, we assume that the system is in the diffusive limit[11] and make use of the

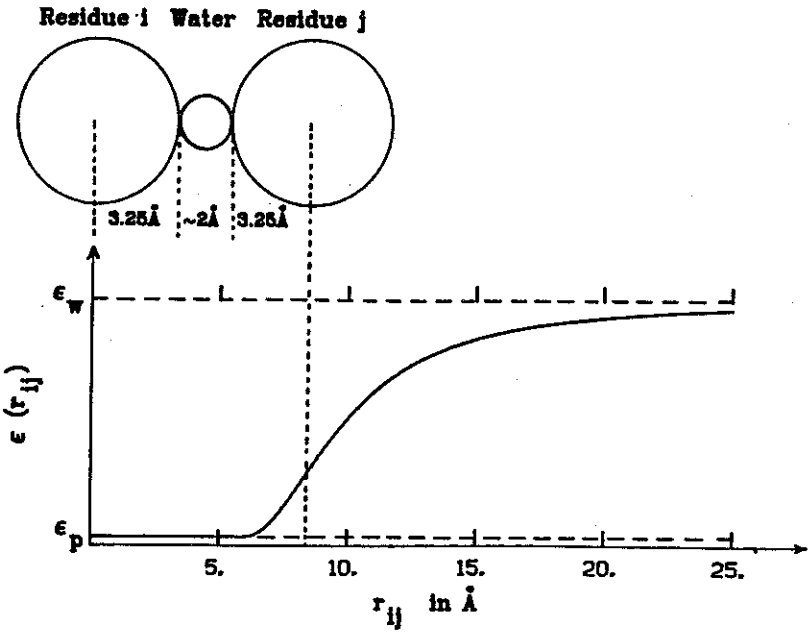


Fig. 1. Form of the dielectric function $\epsilon(r_{ij})$ [see Eq. (11)]. $\epsilon_w = 81$ and $\epsilon_p = 2.5$.

corresponding form of the Langevin equations of motion:

$$\frac{d\mathbf{r}_i}{dt} = -\frac{D}{k_B T}\frac{\partial V}{\partial \mathbf{r}_i} + \frac{D}{k_B T}\mathbf{R}_i \tag{12}$$

Here k_B is the Boltzmann's constant and T is the absolute temperature of the medium. D is the diffusion coefficient; the value of D used for the valine residues is $D = 67.1$ Å^2/ns, which corresponds to a sphere of radius 3.25 Å in water at 25°C. The position vector of the ith residue is denoted by $\mathbf{r}_i$. $\mathbf{R}_i$ represents the randomly fluctuating force exerted on the ith residue by the surrounding fluid, and it is assumed to be a Gaussian random variable with mean and covariance

$$\langle \mathbf{R}_i(t) \rangle = \mathbf{0} \tag{13}$$

$$\langle \mathbf{R}_i(t)\tilde{\mathbf{R}}_j(t') \rangle = 2k_B T\xi \underset{\sim}{1}\delta_{ij}\delta(t - t') \tag{14}$$

where ξ is the friction coefficient that is related to D by the Stokes–Einstein relation, $D = k_B T/\xi$, $\tilde{\mathbf{R}}_j$ denotes the transpose of $\mathbf{R}_j$, $\mathbf{R}_i\tilde{\mathbf{R}}_j$ represents a dyad product, and $\underset{\sim}{1}$ is the unit tensor.

The Langevin equations in Eq. (12) are integrated by using the algorithm of Ermak and McCammon.[18] This yields the position of the ith residue at time $t + \Delta t$ as

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) - \frac{D}{k_B T}\frac{\partial}{\partial \mathbf{r}_i}V[\mathbf{r}_i(t)]\Delta t + \chi_i \tag{15}$$

where the quantity χ_i, which is the displacement due to the stochastic force,

is sampled from a Gaussian distribution with zero mean and width

$$\langle \chi_i \tilde{\chi}_i \rangle = 2D\Delta t \underset{\sim}{1} \tag{16}$$

The time-step size Δt is limited by the condition that the systematic force does not change appreciably during Δt. We use a variable time-step size algorithm[19] to reduce the numerical error due to abrupt changes in the systematic force during a step. In this algorithm, when the magnitude of the change in potential energy during a step exceeds a chosen value, ΔV_{max}, the step is rejected and reexecuted with an adjusted (smaller) time-step given by

$$\Delta t' = \Delta t / \{ [|\Delta V(\text{rejected step})| / \Delta V_{max}] + 1 \} \tag{17}$$

where the brackets define the largest integer whose magnitude does not exceed the magnitude of the quantity inside. After the rejected step has been taken with the smaller time step, $\Delta t'$, the step size is reset to Δt. The default size of the time-step Δt used in the simulations is 0.04 ps. The value of ΔV_{max} is set to 10 kcal/mol, which is approximately equal to $3[\overline{(\Delta V)^2}]^{1/2}$: the value of $\overline{(\Delta V)^2}$ was estimated from a short test simulation. In the full simulations, we neglect hydrodynamic interactions between the residues, as indicated in Eqs. (12) and (15). The magnitude of the effect of hydrodynamic interaction is estimated in the results and discussion section by combining an analytic method and simulation results.

Each trajectory of the polypeptide chain is started from a conformation having a fully extended random-coil part and two helical microdomains already present at the two ends (see Fig. 2). Each of the helices is made up of eight residues, leaving eight residues in the intervening random-coil part. Although all the dihedral angles are allowed to vary throughout the simulation, the helical microdomains retain their overall shapes due to the helix-stabilization energy term V_α [see Eq. (4)]. The above choice of the initial conformation is arbitrary. However, it is shown that the results obtained do not depend on the specific choice of initial conformation.

Twelve trajectories of varying lengths (see footnote of Table I) were generated for a total time of 819.2 ns. The exact length of a given trajectory does not matter because the parameters are chosen such that a stable folded structure does not form. Instead, the system folds and unfolds many times. It is from these results that the rate constants are determined (see the next section). In fact, the unfolding rate is found to be greater than the folding rate for the present model polypeptide (see the results and discussion section). To investigate the role of the attractive hydrophobic interaction, we generate three additional trajectories for chains excluding the attractive interaction term V_{sol} in Eq. (8). The total length of these three trajectories was 134.4 ns (see footnote of Table III).

Analysis of Brownian Dynamics Trajectory

In this subsection, we describe the procedure used for extracting the kinetic information from the Brownian dynamics trajectories. Although we employ

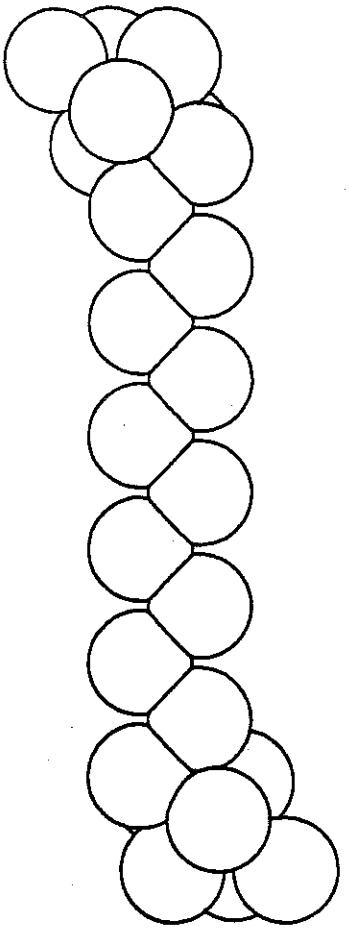


Fig. 2. Initial conformation of the model polypeptide chain. The first eight residues and the last eight residues form the α-helical structures. The initial values of the dihedral angles are 38.3° for ϕ_i ($i = 4, 5, 6, 7, 8, 20, 21, 22, 23, 24$), 110° for ϕ_j ($j = 9, 19$), and 180° for ϕ_k ($k = 10, 11, \ldots, 18$).

the language of the protein folding problem, the methodology is applicable to other problems involving intrachain reactions.

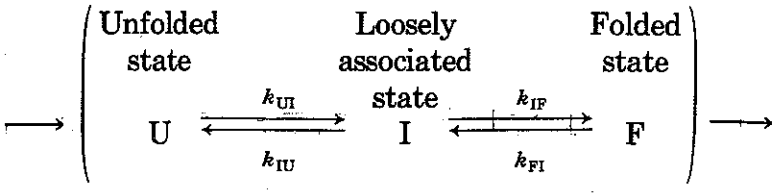
Our object is to evaluate the rate constants k_f and k_r in the following kinetic scheme:

$$\longrightarrow \left(\begin{array}{ccc} \text{Unfolded state} & & \text{Folded state} \\ \text{U} & \underset{k_r}{\overset{k_f}{\rightleftharpoons}} & \text{F} \end{array} \right) \longrightarrow \qquad (18)$$

In general, the "folded state" (F) may be an intermediate state in the overall folding process, and the unfolded state (U) is one representing an earlier intermediate where the two microdomains under consideration have not yet coalesced. In many cases, F is best characterized by the proximity of a specific residue (or residues) of one microdomain and a specific residue (or residues) of the other. The unfolded state corresponds to the extended conformations where short-range interactions, such as the hydrophobic and van der Waals terms, between the two microdomains are essentially zero. For the present

calculation, the folded state corresponds to having the centers of the two helices less than 16 Å apart (see below).

To proceed with the analysis, it is necessary to define a boundary between the folded and unfolded states. This is done most conveniently in terms of the kinetic scheme



(19)

where the "loosely associated state" (I) corresponds to a very narrow zone in the conformational space between U and F. Equation (19) is the usual kinetic scheme for solution reactions with I representing the encounter complex. By applying the steady-state approximation for the population of I, the rate constants obtained from this scheme can be connected with k_f and k_r:

$$k_f = k_{UI}k_{IF}/(k_{IU} + k_{IF}) \tag{20}$$

$$k_r = k_{FI}k_{IU}/(k_{IU} + k_{IF}) \tag{21}$$

In Eq. (20), k_{UI} is the rate constant governing the rate at which the polypeptide chain, starting from the unfolded conformations, attains a conformation I, and the factor $k_{IF}/(k_{IU} + k_{IF})$ gives the probability that it will subsequently attain the correctly folded conformation rather than going back to the unfolded conformations. Equation (21) has a corresponding interpretation.

The rate constant k_{KL} (KL = UI, IU, IF or FI) in the kinetic scheme in Eq. (19) is given by the inverse of the mean first-passage time τ_{KL} from state K to state L.[20] First-passage times, from state K to state L, can be determined from a trajectory in the following way: Let us first consider a first-passage time from I to F. Operationally, we introduce a "clock" and follow the trajectory record as a function of time (see Fig. 3). We set the clock to zero as soon as the trajectory crosses the boundary I from a U-conformation to an F-conformation. The trajectory may meander along the surface region of state F for a while, as depicted in Fig. 3. Up to this point we cannot say that the trajectory has passed into state F, because it could rapidly recross the boundary region I and go back to state U. If we counted all of the crossings of the boundary I as passages into F or U, the transition rates between U and F would be overestimated.

A corresponding problem arises in the calculation of dihedral angle transition rates from Brownian dynamics trajectories,[21,22] where the transitions between the *trans* and *gauche* states are defined, not in terms of barrier crossings but rather as the passage from one potential minimum to another minimum. Here, we define the passage into state F as the arrival at some typical conformation in the central region of F, as represented by the dashed line in Fig. 3. When this event occurs, we record the time as a first-passage time from I to F and reset the clock to zero. After spending some time in F, the trajectory will recross the boundary I in an attempt to return to state U. We record the time as a first-passage time from F to I and reset the clock to zero. The first-passage times from I to U and from U to I are similarly

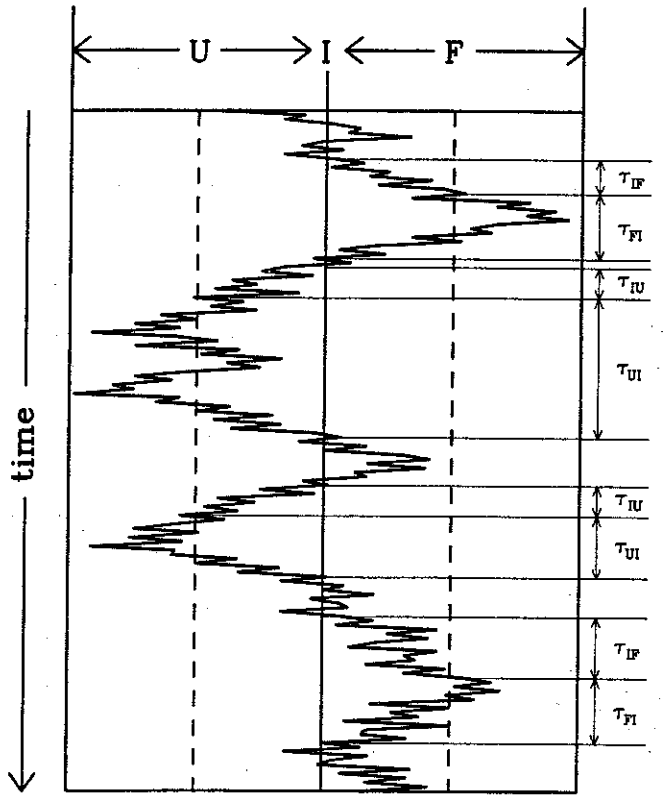


Fig. 3. Plot of a trajectory history. Although we depict the diagram such that the conformational states may be determined by the value of a single gross conformational variable (e.g., the separation between the microdomains), the abscissa must be interpreted in an abstract way. In general, it represents a collection of variables that defines the conformational states U, I, and F. The dashed lines locate the typical conformations of central regions of U and F. For the description on the first-passage times τ_{ij}, see text.

obtained. It should be noted that only direct transitions from I to U or F are included; i.e., the clock is set to zero whenever the trajectory crosses the boundary I in either direction. This is because a trajectory that sets forth from I and is eventually to pass into F (U) may spend some time in the surface region of U (F). This period of time is not the one spent in going from I to F (U) and, therefore, must not be included in the calculation of the first-passage time from I to F (U) (see Fig. 3).

To apply the formulation just outlined to the present model problem, it is necessary to introduce specific operational definitions for the various states in the kinetic scheme (Eq. (19)]. The interaction energy of the two helices, with the neglect of the contribution from the intervening random-coil part, is plotted in Fig. 4(a) as a function of the six variables shown in Fig. 5. We estimate the orientation-averaged interaction energy as a function of r (the distance between the centers of the two helices; see Fig. 5) by the method of Koide and Kihara.[23]* The results are shown in Fig. 4(b). Since the interaction

*The orientationally averaged interaction potential energy given in Fig. 4(b) was calculated approximately by using the procedure proposed by Koide and Kihara.[23] Strictly speaking, their method is applicable only to cylindrically symmetric molecules with a mirror plane perpendicular to the symmetry axis such as N_2, O_2, and CO_2. The helices in the present problem may be considered to be roughly cylindrical, in particular for $r > 16$ Å. Due to the presence of helix dipoles [see Eq. (9)], however, the helix lacks the said mirror plane. Hence Fig. 4(b) represents only an estimate that is sufficiently accurate for the present purpose.

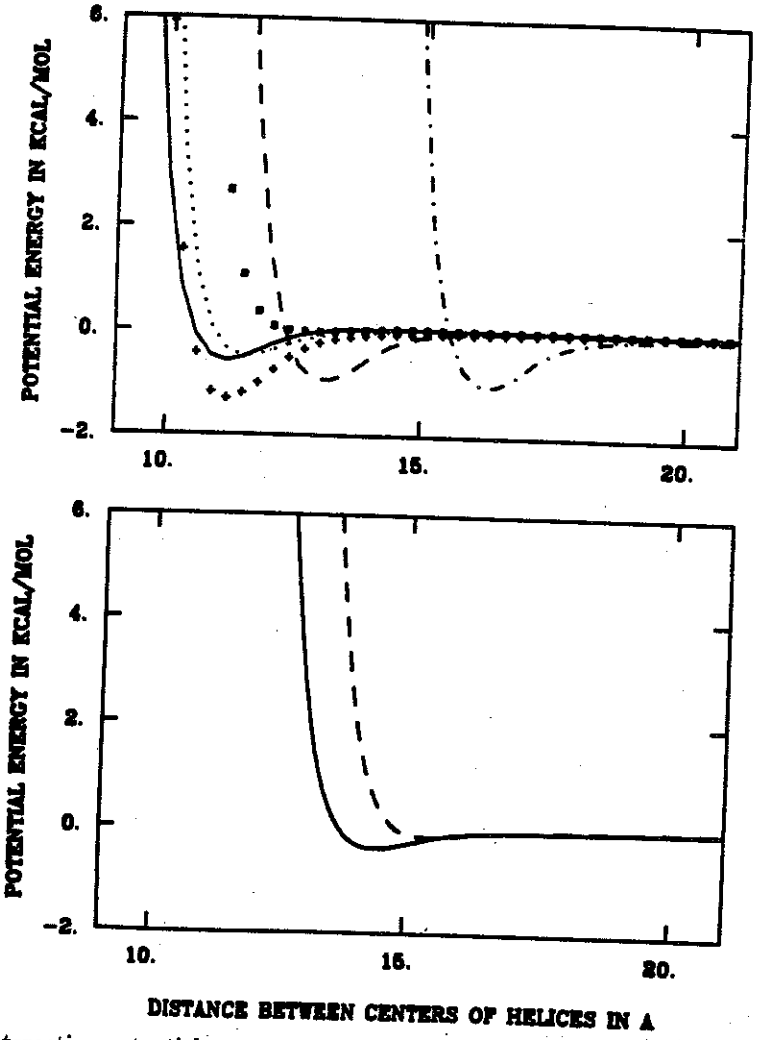


Fig. 4. (a) Interaction potential energy between the helices at some typical orientations. The definitions of orientational angles, α, β_1, β_2, γ_1, and γ_2 are given in Fig. 5. (—) $\alpha = 0°$, $\beta_1 = 90°$, $\beta_2 = 90°$, $\gamma_1 = 0°$, $\gamma_2 = 0°$. (---) $\alpha = 0°$, $\beta_1 = 0°$, $\beta_2 = 90°$, $\gamma_1 = 0°$, $\gamma_2 = 0°$. (·····) $\alpha = 0°$, $\beta_1 = 0°$, $\beta_2 = 0°$, $\gamma_1 = 0°$, $\gamma_2 = 0°$. (· · · ·) $\alpha = 90°$, $\beta_1 = 90°$, $\beta_2 = 90°$, $\gamma_1 = 0°$, $\gamma_2 = 0°$. (+ + +) $\alpha = 180°$, $\beta_1 = 90°$, $\beta_2 = 90°$, $\gamma_1 = 0°$, $\gamma_2 = 0°$. (□□□) $\alpha = 0°$, $\beta_1 = 90°$, $\beta_2 = 90°$, $\gamma_1 = 180°$, $\gamma_2 = 0°$. (b) Interaction potential energy averaged over β_1 and β_2. (—) $\alpha = 180°$, $\gamma_1 = \gamma_2 = 0°$; (---) $\alpha = 0°$, $\gamma_1 = \gamma_2 = 0°$.

potential energy vanishes for $r \gtrsim 16$ Å, we may define the unfolded state U simply as those conformations where $r > 16$ Å. Accordingly, we define I as the boundary region at $r \approx 16$ Å, which is consistent with a steady-state approximation for its population. Finally, the folded state F corresponds to $r < 16$ Å.

For the reason discussed above, we need to make the operational definitions for the transitions from I to F or U more precise. We will say that a trajectory passes into F if any one or both of the middle residues on one helix come into contact with any one or both of the middle residues on the other helix; *contact* is defined as an interresidue separation less than 6.5 Å (sum of the van der Waals radii of the residues). Similarly, a trajectory is said to pass into U if the interhelical separation exceeds r', the median value for the unfolded state

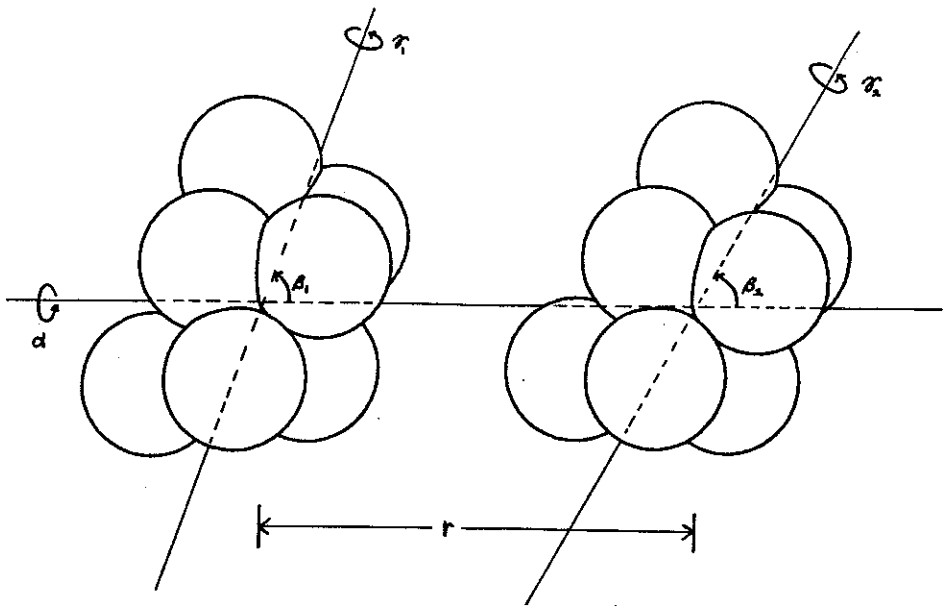


Fig. 5. Coordinates that describe the relative orientation and the separation between the two helices.

population; that is,

$$\int_a^{r'} dr\, P_{eq}(r) = \int_{r'}^{b} dr\, P_{eq}(r) \tag{22}$$

where $P_{eq}(r)$ is the equilibrium probability that the interhelical separation is equal to r (see Fig. 6), b is the maximum possible separation between the centers of helices ($b = 44$ Å), and $a = 16$ Å. Using the calculated value of $P_{eq}(r)$ from the trajectories (see Fig. 6) we find $r' = 26.8$ Å.

ANALYTIC MODEL

In this section, we examine previous analytic treatments of the folding dynamics between two microdomains,[1,2,4-8] and show how to relate the parameters appearing in these treatments to the simulation results.

General Formulation

In the diffusion-collision model for two microdomains, it is assumed that the effects from other parts of the polypeptide chain can be treated implicitly. This assumption reduces the full many-body problem into a two-body problem involving the relative motion of the two microdomains. The Smoluchowski equation for the motions of two microdomains is[24]

$$\frac{\partial}{\partial t} P(\mathbf{r}_1, \Omega_1, \mathbf{r}_2, \Omega_2, t) = -\tilde{\mathbf{M}} \cdot \mathbf{J}(\mathbf{r}_1, \Omega_1, \mathbf{r}_2, \Omega_2, t) \tag{23}$$

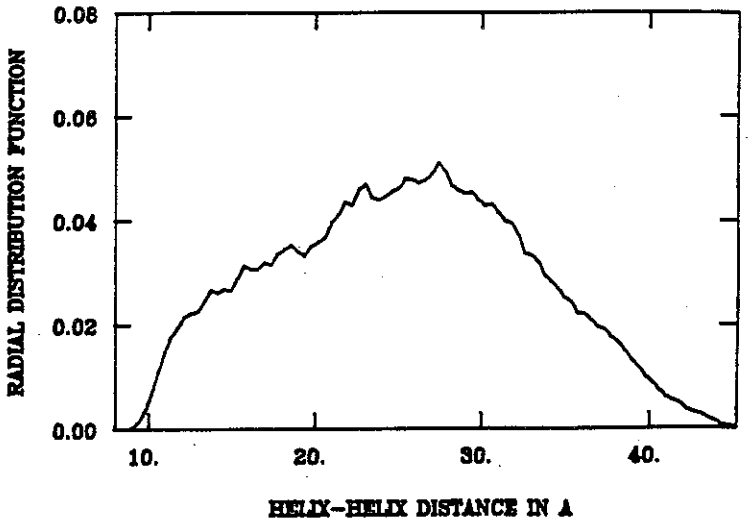


Fig. 6. Radial distribution function $P_{eq}(r)$ calculated from the Brownian dynamics trajectories.

where $P(\mathbf{r}_1, \Omega_1, \mathbf{r}_2, \Omega_2, t)$ is the probability density that the positions and orientations of two microdomains are $(\mathbf{r}_1, \Omega_1)$ and $(\mathbf{r}_2, \Omega_2)$ at time t. The vector operator $\mathbf{M}$ represents the ∇ operation in the 12-dimensional space, and $\tilde{\mathbf{M}}$ denotes the transpose. The explicit expression for $\mathbf{M}$ is quite lengthy and will not be reproduced here (see Ref. 24). The flux $\mathbf{J}$ in the 12-dimensional space is given by

$$\mathbf{J}(\mathbf{r}_1, \Omega_1, \mathbf{r}_2, \Omega_2, t) = -\underset{\sim}{\mathbf{D}}(\mathbf{r}_1, \Omega_1, \mathbf{r}_2, \Omega_2) \cdot \left[\mathbf{M}P(\mathbf{r}_1, \Omega_1, \mathbf{r}_2, \Omega_2, t) + \frac{1}{k_B T} P(\mathbf{r}_1, \Omega_1, \mathbf{r}_2, \Omega_2, t)\mathbf{M}G(\mathbf{r}_1, \Omega_1, \mathbf{r}_2, \Omega_2)\right] \quad (24)$$

where $\underset{\sim}{\mathbf{D}}(\mathbf{r}_1, \Omega_1, \mathbf{r}_2, \Omega_2)$ is the 12-dimensional diffusion tensor, and $G(\mathbf{r}_1, \Omega_1, \mathbf{r}_2, \Omega_2)$ is the potential of mean force for the microdomain motions. G differs from the bare interaction potential energy between the microdomains due to the presence of other parts of the polypeptide chain and the solvent modification of the potential. The "neglected" parts can make an entropic contribution that favors a collapsed relative to an extended chain. Moreover, the neglected parts resist changes in the overall conformation (e.g., relative separation between the two microdomains under consideration) due to potential barriers for bond rotation, steric hindrance, etc., so that the diffusion tensor $\underset{\sim}{\mathbf{D}}$ is different from that for the microdomains connected by a featureless string. By use of the simulation results, we are able to evaluate the magnitude of these effects (see the results and discussion section).

Analytic solution of Eqs. (23) and (24) for the general case involving nonspherical microdomains whose folding probability is orientationally restricted is not possible. To simplify the problem, it is assumed that the probability density function P and the potential of mean force G are independent of the overall position and orientation of the entire peptide chain, and that their dependence on the relative orientation of the two microdomains may be spherically averaged. For the spherically averaged probability density function $p_r(r, t)$ and potential of mean force $g(r)$, we have[8]

$$\frac{\partial}{\partial t}p_r(r, t) = -\frac{1}{r^2}\frac{\partial}{\partial r}[r^2 j_r(r, t)] \tag{25}$$

$$j_r(r, t) = -D(r)\left[\frac{\partial P_r(r, t)}{\partial r} + \frac{1}{k_B T}p_r(r, t)\frac{\partial g(r)}{\partial r}\right] \tag{26}$$

Equations (25) and (26) are solved with the following boundary conditions at $r = a$, where contact between the two microdomains occurs and at $r = b$ which is the maximum possible separation:

$$j_r(a, t) = -D(a)\frac{\beta}{l\gamma}p_r(a, t) \qquad \text{at } r = a \tag{27}$$

and

$$j_r(b, t) = 0 \qquad \text{at } r = b \tag{28}$$

In Eq. (27), β is the probability of reaction ($0 \leq \beta \leq 1$), γ ($\gamma = 1 - \beta$) is the probability of reflection at $r = a$, and l is a characteristic length whose physical meaning was discussed in Refs. 1 and 2. The boundary condition, at $r = b$ given by Eq. (28), corresponds to a totally reflecting boundary. The average dependence of the folding probability on the relative orientation of microdomains and other neglected internal conformational variables is implicitly included in the choice of values for the set of parameters, β, γ, and l. These are the parameters that will be related to the rate constants defined in the kinetic scheme in Eq. (19) and evaluated by the Brownian dynamics simulation.

Mean Coalescence Time

The exact solution to Eqs. (25)–(28) cannot be given in the closed form[1,2] for a general potential of mean force. However, by assuming a simple exponential decay for the probability $\Sigma(t)$ that the microdomains are still uncoalesced at time t [$\Sigma(t) = \int_a^b dr\, 4\pi r^2 p_r(r, t)$],

$$\Sigma(t) \approx \Sigma_{\text{approx}}(t) = \exp(-t/\tau) \tag{29}$$

we can obtain a simple general expression for the mean coalescence time τ, which is the reciprocal of the rate constant k_f;[8] the approximation given by Eq. (29) has been shown to be valid in many cases.[8] For the initial condition

that the radial separation is r_0 at $t = 0$, the result is

$$\tau(r_0) = \int_a^{r_0} dr \frac{e^{g(r)/k_BT}}{r^2 D(r)} \int_r^b dq\, q^2 e^{-g(q)/k_BT} + \frac{l\gamma}{\beta D(a)} \frac{e^{g(a)/k_BT}}{a^2} \int_a^b dr\, r^2 e^{-g(r)/k_BT} \tag{30}$$

For an equilibrium initial distribution,

$$p_r(r,0) = r^2 e^{-g(r)/k_BT} \Big/ \int_a^b dr\, r^2 e^{-g(r)/k_BT} = P_{eq}(r) \tag{31}$$

we obtain[8]

$$\tau = \int_a^b dr \left[D(r) P_{eq}(r)\right]^{-1} \left[\int_r^b dq\, P_{eq}(q)\right]^2 + \left[\frac{\beta D(a)}{l\gamma} P_{eq}(a)\right]^{-1} \tag{32}$$

The simplest application of the analytic model to microdomain coalescence is that given previously.[1,2] It was assumed there that the two microdomains are connected by a featureless string and that the interaction energy is short ranged. Then, the coalescence dynamics between the two microdomains is determined essentially by the boundary conditions. Assuming that $g(r) = 0$ for $r > a$ and setting $D(r)$ equal to a constant value $D_1 + D_2$ (D_α is the diffusion coefficient of microdomain α), we obtain from Eq. (30)

$$\tau(r_0) = \frac{1}{3(D_1 + D_2)} \left[b^3 \left(\frac{1}{a} - \frac{1}{r_0}\right) - \tfrac{1}{2}\left(r_0^2 - a^2\right)\right] + \frac{l\gamma}{\beta} \frac{1}{3(D_1 + D_2)} \frac{1}{a^2} (b^3 - a^3) \tag{33}$$

A slightly improved treatment can be obtained by assuming a square potential well for $g(r)$:

$$g(r) = \begin{cases} -V_0 & \text{if } r \le r_1 \\ 0 & \text{if } r > r_1 \end{cases}$$

Assuming again that $D(r) = D_1 + D_2$, we obtain

$$\tau(r_0) = \frac{1}{3(D_1 + D_2)} \left[e^{-V_0/k_BT}\left(b^3 - r_1^3\right)\left(\frac{1}{a} - \frac{1}{r_1}\right) - \tfrac{1}{2}\left(r_0^2 - a^2\right) + b^3\left(\frac{1}{r_1} - \frac{1}{r_0}\right) + r_1^3\left(\frac{1}{a} - \frac{1}{r_1}\right)\right] + \frac{l\gamma}{\beta} \frac{1}{3(D_1 + D_2)} \frac{1}{a^2} \left[\left(r_1^3 - a^3\right) + e^{-V_0/k_BT}\left(b^3 - r_1^3\right)\right] \tag{34}$$

The validity of the approximate results given in Eqs. (33) and (34) are examined in the results and discussion section.

Evaluation of Model Parameters from Simulation Results

Although the parameters β, γ, and l in the analytic treatment can be interpreted in terms of theoretical models,[1,2] it is possible to relate them to the kinetic scheme [Eq. (19)] and evaluate them from the Brownian dynamics simulation. We rewrite Eq. (20) as

$$\frac{1}{k_f} = \frac{1}{k_{UI}} + \frac{1}{k_{IF}} \frac{k_{IU}}{k_{UI}} \tag{35}$$

As discussed above, $1/k_{UI}$ is the mean first-passage time τ_{UI} from U to I. In the calculation of τ_{UI}, events that follow the arrival at I do not matter. Thus τ_{UI} corresponds to the first term on the right-hand side of Eq. (30), which is the mean coalescence time obtained with the completely absorbing boundary condition at $r = a$ (i.e., $\gamma/\beta = 0$).[2,8] Accordingly, the second term in Eq. (35) corresponds to the second term in Eq. (30). It represents the mean time τ_{IF}^* required for attaining the correctly folded state F after starting from the loosely associated state at $r = a$. This mean time is different from τ_{IF} defined above, because τ_{IF}^* includes the time spent in the unfolded state U after first crossing the boundary;[25] i.e., $\tau_{IF}^* = \tau_{IF}(\tau_{UI}/\tau_{IU})$. By inserting Eq. (31) into the second term of Eq. (30), the latter can be written as $(l\gamma/\beta)[D(a) \cdot P_{eq}(a)]^{-1}$. Comparing this result with Eq. (35), we have

$$\frac{l\gamma}{\beta} = D(a)P_{eq}(a)\frac{1}{k_{IF}}\frac{k_{IU}}{k_{UI}} \tag{36}$$

RESULTS AND DISCUSSION

Here we present the simulation results for the folding and unfolding rates of the model helical polypeptide chain, and use them to examine several important aspects of the analytic model described in the previous section. They include the nature of the potential of mean force between the microdomains $g(r)$ and the position-dependent effective diffusion coefficient $D(r)$. Also, we determine the role of the short-ranged attractive interaction of the hydrophobic type and estimate the magnitude of the effect of hydrodynamic interaction.

Rate Constants for the Folding and Unfolding of the Model Polypeptide

To calculate the mean first-passage times τ_{UI}, τ_{IU}, τ_{IF}, and τ_{FI} from Brownian dynamics trajectories, it is necessary to perform the analysis outlined above. Figure 7 shows the variation of r with time for all trajectories with the full potential. Table I lists the first-passage times found from the trajectories. Each of the various rate constants in the kinetic scheme in Eq.

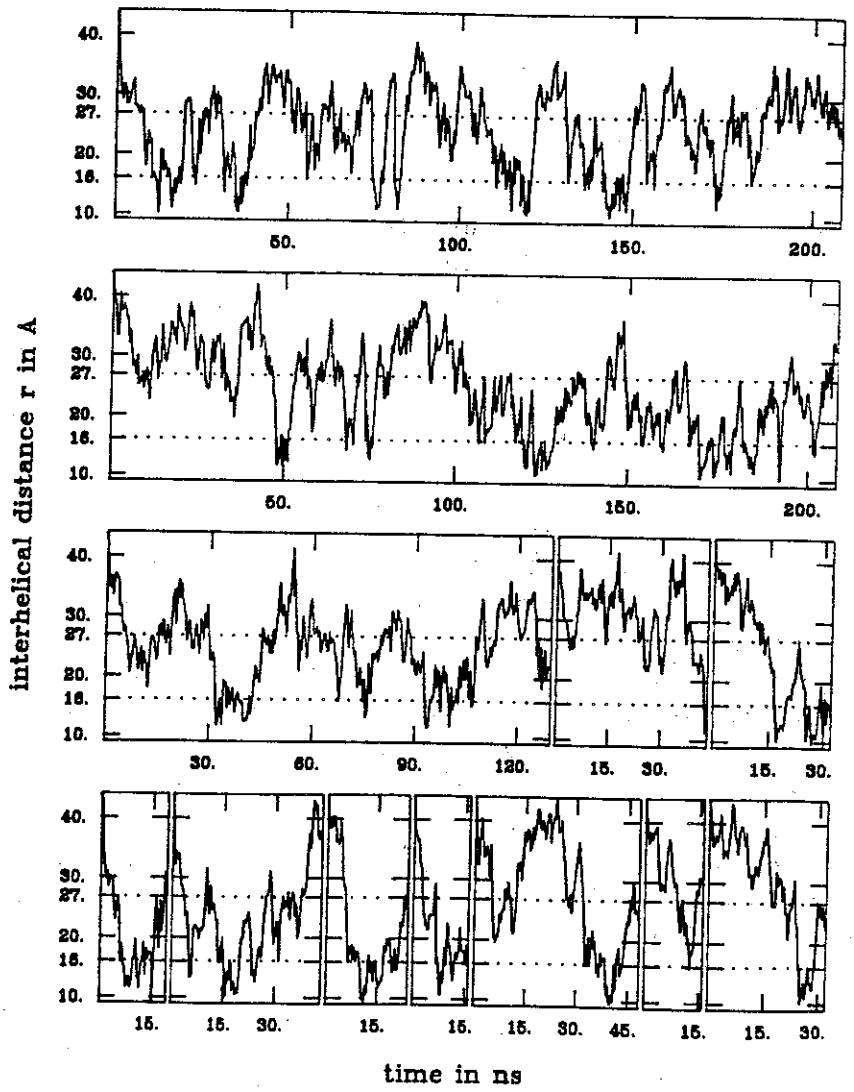


Fig. 7. Variation of the distance between centers of helices with time obtained from all of the normal chain trajectories. Dotted lines represent the separations, $r' = 26.8$ Å and $a = 16$ Å.

(19) are given by the inverse of the corresponding mean first-passage time, and the folding and unfolding rate constants k_f and k_r are calculated from Eqs. (20) and (21). The results are presented in Table II, together with the estimates of the statistical errors (see the appendix).

Role of the Short-Ranged Attractive Interaction

In Table III, we list the observed first-passage times between states U, I, and F for the chains without hydrophobic attraction. In Fig. 8, we plot the variation of r with time for all three trajectories of the chain made up of "nonsticky" residues along with that for normal chain trajectories generated by using almost the same sequences of random number seeds for obtaining the random displacement χ_i [see Eq. (15)]; they are not exactly the same because of the variable step size used. It can be seen that the general behavior of each trajectory of the nonsticky chain is very similar to that of the corresponding normal chain trajectory. However, for these shorter trajectories, only one

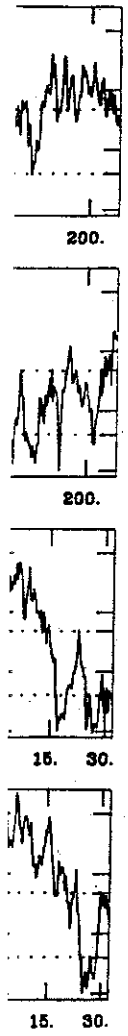


TABLE I
Observed First-Passage Times (in ps).[a]

Observed first-passage times from U to I	3045.0	6240.0	14997.0	18081.0	2517.0	27023.0
	9763.0	2870.0	2693.0	3879.0	14885.0	7188.0
	14988.0	2736.0	30090.0	511.0	6256.0	5627.0
	9918.0	5061.0	1088.0	1936.0	7308.0	22114.8
	5940.4	13030.8	1984.0	3834.8	3837.2	
Mean first-passage time τ_{UI}: 8601.45 ps						
Observed first-passage times from I to U	702.0	1730.0	1808.0	1057.0	759.0	905.0
	803.0	1617.0	378.0	1213.0	1701.0	1373.0
	2053.0	1115.0	929.0	960.0	1195.0	1320.0
	3044.0	754.0	2121.0	866.0	3577.0	1431.0
	1537.0	2534.8	1203.2	3815.6	367.2	1855.0
	643.2	3550.4	916.4	782.4	1534.0	1429.0
	1349.0	648.0				
Mean first-passage time τ_{IU}: 1462.53 ps						
Observed first-passage times from I to F	572.0	197.0	202.0	315.0	91.0	532.0
	139.0	848.0	265.0	115.0	128.0	248.0
	188.0	162.0	456.0	326.0	301.6	401.0
	205.0	397.0	496.0	404.4	121.6	161.6
	496.0	528.0	519.2	196.4	441.0	116.0
	324.0					
Mean first-passage time τ_{IF}: 319.12 ps						
Observed first-passage times from F to I	451.0	1462.0	703.0	1192.0	453.0	1029.0
	2708.0	505.0	439.0	343.0	104.0	818.0
	2052.0	905.0	2045.0	1376.0	994.0	2397.0
	870.0	1899.0	198.0	1909.2	1384.4	327.6
	1499.2	2287.2	1936.8	1850.8	2160.0	463.0
	1951.0					
Mean first-passage time τ_{FI}: 1248.78 ps						

[a] Results from 12 trajectories of length 208.0, 208.0, 129.6, 43.2, 32.0, 19.2, 43.2, 24.0, 16.0, 48.0, 16.0, 32.0 ns.

TABLE II
Rate Constants for the Diffusion-Collision Step Between Two α-Helical Microdomains at Ends of a Polypeptide Chain

	U → I	I → U	I → F	F → I
Number of passages observed	29	38	31	31
Error limits $\binom{x}{x'}$ with $2\alpha = 0.1$[a]	20.1	27.8	21.8	21.8
	39.0	49.3	41.3	41.3
Mean first-passage time (in ns)	8.60	1.46	0.32	1.25
Minimum first-passage time	0.51	0.37	0.09	0.10
Maximum first-passage time	30.09	3.82	0.85	2.71
Rate constant k_{KL} (in ns^{-1})	0.12	0.68	3.13	0.80
Error limits $\binom{x}{x'}$ with $2\alpha = 0.1$[a]	0.08	0.50	2.21	0.56
	0.16	0.89	4.18	1.07
k_f (in ns^{-1})	0.096			
k_r (in ns^{-1})	0.14			

[a] See in the appendix.

TABLE III
Observed First-Passage Times (ps) for Chains Without the Attractive Hydrophobic Term in Eq. (8)[a]

Observed first-passage times from U to I: 24642.0 3436.0 39432.0 900.0 2024.0 7244.0 5834.0
τ_{UI} = 11930 ps
k_{UI} = 0.084 ns^{-1}

Observed first-passage times from I to U: 1444.0 726.0 2810.0 920.0 726.0
τ_{IU} = 1325 ps
k_{IU} = 0.75 ns^{-1}

Observed first-passage times from I to F: 370.0
Observed first-passage times from F to I: 92.0

[a] Results from three trajectories of length 38.4, 48.0, 48.0 ns.

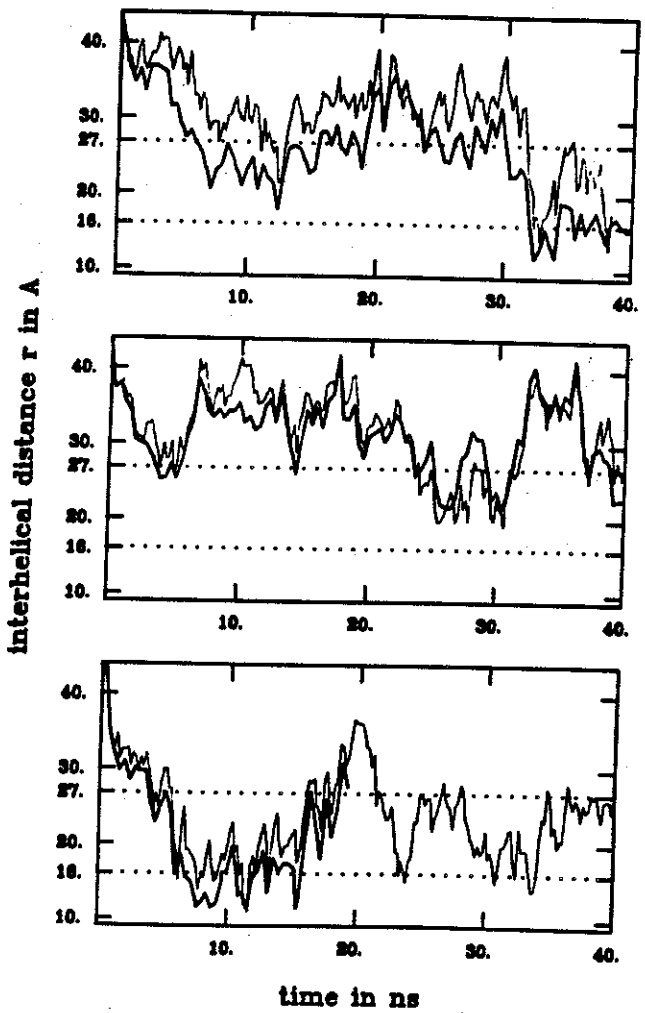


Fig. 8. Variation of the interhelical distance r with time. Thick solid curves are for the normal chains and thin solid curves are for the chains without hydrophobic attraction (see text).

transition to and from the folded state was observed (see Table III). For τ_{UI} and τ_{IU} the results are similar to those in the presence of hydrophobic interaction, although there is a large uncertainty due to the limited statistics. Corresponding calculations with and without the electrostatic interaction show that it has little effect on the folding rate.

Potential of Mean Force $g(r)$

In Refs. 1 and 2, it was argued that an important force between the uncharged microdomains is of the hydrophobic type. This is expected to be relatively short ranged; that is, until the two microdomains are sufficiently close to exclude water molecules, no hydrophobic attraction exists. Consequently, with the intervening chain segment defined as a featureless string, it was supposed that most of the relative motion of the microdomains would involve essentially free diffusion. Such an assumption appears applicable to the present model, since the potential between the helices (excluding the contribution from the intervening chain segment) is flat through most of its range (see Fig. 4).

The simple model in which the two helices are connected by a featureless string does not take into account the effects of the intervening peptide chain. One of these is the deviation of the equilibrium probability distribution $P_{eq}(r)$ from a simple r^2 dependence expected for $r > 16$ Å; for $r \lesssim 16$ Å, the attractive interaction would increase the population [cf. Eq. (31)]. Also, as discussed subsequently, the chain alters the helix diffusion coefficient. In Fig. 6, we see that, contrary to expectation from the simple model, $P_{eq}(r)$ is peaked near 27 Å and decreases with increasing distance. This is due to the entropic contribution of the intervening chain to the effective potential of mean force between the helices. The separation between the helices is determined by the conformational state of the intervening chain. There are more conformational states that give the intermediate separation in the neighborhood of 27 Å than those available for the fully extended geometry. A similar trend has been observed by Haas et al.[26] in a study of the end-to-end distance in a series of oligopeptides.

The form of $g(r)$ governing the relative motion of the helices is obtained from the equation

$$g(r) \sim -k_B T \ln\left[P_{eq}(r)/r^2\right] \tag{37}$$

The result is shown in Fig. 9. The repulsive part of the potential (cf. Fig. 4) is dominant at shorter distances and the minimum occurs at $r \approx 12$ Å due to the attractive interaction that favors the folded conformation. In the region of constant potential, the entropic effect from the intervening chain gradually increases $g(r)$ as r increases. Finally, for large $r(r \to b)$, there are very few conformations available for the chain and $g(r)$ rises rapidly.

The effect of the intervening chain on the relative motion of microdomains is likely to be important in the folding of real polypeptide chains. By reducing the possible conformations available for a polypeptide chain to those in which

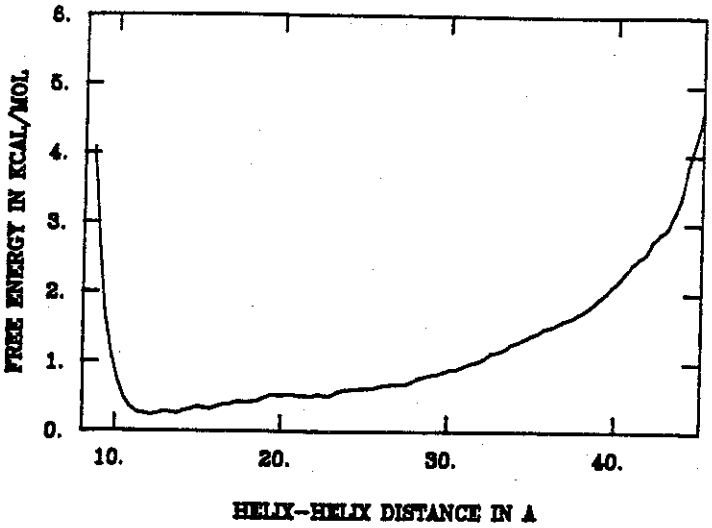


Fig. 9. Interaction free energy $g(r)$ calculated by using Eq. (37).

the chain is not fully extended, the time required to form the native structure, which is very compact in globular proteins, may be reduced.

Effective Diffusion Coefficient

The relative diffusion coefficient $D(r)$ appearing in Eqs. (30) and (32) is not simply the sum of the translational diffusion coefficients of the free microdomains. It depends on the dynamic effect of the neglected part of the polypeptide chain on the motion of the microdomains. The interhelical distance r can change only if there are internal rotations in the random-coil part. Since there are potential energy barriers to these internal rotations, we may expect that the relative diffusional motion of the helices is retarded by the presence of the intervening chain, as well as by the friction from the viscous medium; a detailed analysis of this effect will be given separately. As a first approximation, we use the concept of internal viscosity[27,28], η_{int}, that is taken to vary as $\eta_{int} = A + B\eta_0$. Here the first term A accounts for the effect of the intervening chain and becomes the dominant term as the solvent viscosity η_0 goes to zero. Since A and B are complicated functionals of temperature, ionic strength, and molecular properties of the chain and solvent,[27,28] analytic evaluation of A and B is difficult. Hence, in previous model studies of the protein folding,[1-9] it was simply assumed that $\eta_{int} \approx \eta_0$. This assumption is valid only if the solvent is very viscous and the microdomains are large in size compared to the neglected part of the chain.

To estimate the ratio η_{int}/η_0 for the present system, we compare the translational diffusion coefficient for the free helix with that obtained from the simulation for the bound helix. For a rigid helix made up of N residues with stick boundary condition, the translational diffusion coefficient D_F is given by[29-31]

$$D_F = (k_B T)/(6\pi N \eta_0 \sigma) \tag{38}$$

in the free-draining limit used for the Brownian dynamics simulation; here σ

is the hydrodynamic radius of a residue. The value of D_F obtained for the present case ($N = 8$, $\sigma = 3.25$ Å) is 8.4 Å^2/ns.

The apparent translational diffusion coefficient of a bound helix, D_E, can be calculated in the standard way from the equation[32]

$$D_E = \frac{1}{6t}\overline{[\mathbf{r}_E(t) - \mathbf{r}_E(0)]^2} \tag{39}$$

where $\mathbf{r}_E(t)$ is the position vector of center of mass of the helix and the bar means a time-average evaluated from the trajectory. To avoid any bias resulting from the effects of $g(r)$ (see Fig. 9), the time average is taken over segments of trajectory where $g(r)$ is almost flat (in the neighborhood of $r \approx 24$ Å). The result is $D_E = 6.7 \pm 0.2$ Å^2/ns, where the estimate of the statistical error in the value of D_E was obtained from an accurate error formula for the translational diffusion coefficient obtained from Eq. (39).[33]

These results can be used to determine η_{int}/η_0 by assuming the Stokes–Einstein formula[27,28] for the relation between the apparent translational diffusion coefficient of a bound helix D_E and the apparent internal viscosity η_{int} by writing

$$D_E = (k_B T)/(C\eta_{int}\sigma) \tag{40}$$

Although C is a quantity depending on a variety of factors, we assume it has the same value as for a free helix [i.e., $C = 6\pi N$ by comparison with Eq. (38)] and obtain $\eta_{int}/\eta_0 = D_F/D_E = 1.25$.

To further examine the effect of the intervening chain length on the effective diffusion coefficient D_E, and consequently on the apparent internal viscosity, we have done a set of calculations with different numbers of residues in the random-coil part of the polypeptide chain. The results are given in Table IV. In all cases, the value of η_0 used was 0.01 g/s · cm, corresponding to diffusion in H_2O at room temperature, and the number of residues in a helix is eight. Hence, the corresponding free-helix diffusion coefficient is 8.4 Å^2/ns. As can be seen from the table for the range of chain lengths considered, D_E and η_{int} are the same within the uncertainty of the calculation.

Evaluation of the Analytic Model

The detailed results obtained from the Brownian dynamics simulation provide information that can be used in conjunction with the analytic model.

Although various theoretical arguments have been given to estimate the parameters in the factor $(l\gamma/\beta)$ appearing in the analytic model (see above), no straightforward procedure for obtaining them is available. As we have pointed out, Eq. (36) enables us to evaluate $l\gamma/\beta$ from the simulation results. The values of the rate constants k_{IF}, k_{IU}, and k_{UI} are given in Table II, and the values of $P_{eq}(a)$ is estimated to be 0.036 Å^{-1} for $a = 16$ Å (see Fig. 6). Then, using the value of $2D_E$ for $D(a)$, we find that $l\gamma/\beta = 0.9$ Å.

To evaluate the effect of the presence of a realistic intervening chain, we compare the folding rate constant k_f obtained from the Brownian dynamics simulation with that calculated from the simple analytic model. In the latter, we use the square-well model for $g(r)$ [see Eq. (34)] and the free-helix

TABLE IV
Apparent Diffusion Coefficients of Helices Bound at the Ends of Polypeptide Chains with Intervening Random-Coil Parts of Varying Lengths[a]

Number residues in random-coil part	Diffusion coefficient of bound helix D_E	η_{int}/η_0
4	6.9 ± 0.5	1.22
6	7.0 ± 0.5	1.21
8	6.7 ± 0.2	1.25
10	7.0 ± 0.3	1.20

[a] Calculated from Brownian dynamics trajectories by use of Eq. (39).

TABLE V
Folding Rate Constants k_f Observed from Brownian Dynamics (BD) Trajectories and Calculated From Eqs. (33), (34), and (30) Without and With the Hydrodynamic Interaction (HI)

	τ_{UI}[a] (ns)	τ_{IF}^*[b] (ns)	k_f[c] (ns^{-1})
Observed value from BD trajectories	8.6	1.9	0.096
Calculated values from			
Eq. (30) without HI	14.2	1.9	0.062
Eq. (30) with HI	8.9	1.8	0.093
Eq. (33)	38.0	5.7	0.023
Eq. (34)	35.1	2.8	0.026

[a] See the section analyzing BD trajectories for its definition; this corresponds to the first term in Eqs. (30), (33), and (34).
[b] See the section evaluating model parameters for its definition; this corresponds to the second term in Eqs. (30), (33), and (34).
[c] $k_f = (\tau_{UI} + \tau_{IF}^*)^{-1}$.

diffusion coefficient for D_1 and D_2 [see Eq. (38)], corresponding to the case with a featureless string. Parameters associated with the square-well potential, roughly fitted to Fig. 4(b), are $V_0/k_BT = 0.7$ and $r_1 = 17$ Å. With the values $a = 16$ Å, $b = 44$ Å, $r_0 = 26.8$ Å, $D_1 = D_2 = D_F = 8.4$ Å^2/ns, and the above value for $l\gamma/\beta$, we obtain $k_f = 0.026$ ns^{-1} from Eq. (34) (see Table V). The result obtained with the cruder model having $V_0 = 0$ [cf. Eq. (33)] is very similar ($k_f = 0.023$ ns^{-1}). Both values are about one fourth of the observed folding rate constant ($k_f = 0.096$ ns^{-1}: see Table II). Hence, it is essential to incorporate a realistic intervening chain in any estimate of the folding rate of the two helices. Of primary significance is the effect of the chain on the potential of mean force $g(r)$; the effect on the diffusion coefficient is much less important.

It is of interest to compare the folding rate constant k_f calculated from the refined analytic model [Eq. (30)] by using $g(r)$, D_E, and $l\gamma/\beta$, obtained from the simulation results, with that found directly from the trajectories. If we assume that $D(r) = 2D_E$, we obtain $k_f = 0.062$ ns^{-1} from Eq. (30) (see Table V). The result is rather close to the value $k_f = 0.096$ ns^{-1} obtained from

the trajectories. To investigate the sensitivity of the latter to the transition distance, we reanalyzed the trajectories by choosing $a = 20$ Å; the distance between the centers of two colinear helices in contact is 19 Å, and thus, for $r \geq 20$ Å, the interaction between the helices is approximately spherical. The actual folding rate estimated from the trajectory calculation ($k_f = 0.076$ ns^{-1}) is only slightly larger than the analytic estimate ($k_f = 0.068$ ns^{-1}) obtained from Eq. (30) with $a = 20$ Å. This suggests that the analytic model described in the third section can be a good approximation if the required parameters [e.g., $g(r)$, D_E, $l\gamma/\beta$] are known. Whether the simulation results obtained here can be used more generally will have to be determined.

Estimation of the Effect of Hydrodynamic Interaction

Although we have neglected hydrodynamic interaction (HI) in the present investigation, we can give a rough estimate of its effects. There are essentially two contributions: the first concerns the effect on the mobility of the helical microdomains and the second concerns the relative motion of the two helices, including the intervening chain. When HI is taken into account, the translational diffusion coefficient of a free helix, calculated by using the numerical method described in Refs. 29–31, is 35.2 Å^2/ns, which is four times as large as the free-draining value. If the effect of internal viscosity in the presence of HI were the same as in the absence of HI ($\eta_{int}/\eta_0 = 1.25$), the value of D_E, including the effect of HI and the intervening chain, would be 28.2 Å^2/ns. However, since the increased mobility of the bound helices in the presence of HI induces more rapid changes in the values of the dihedral angles of the intervening chain, the internal viscosity is expected to be slightly larger with HI than without; i.e., the above value for D_E (with HI) is an upper limit. Thus, the effect of HI among residues within each helix is to decrease τ_{UI} by a factor of approximately 4 relative to the free-draining values.

As mentioned above, there is an additional effect of HI; it retards the approach of one helix to the other. In a spherically averaged treatment, this effect can be described approximately by a position-dependent diffusion coefficient[34] of the form

$$D(r) = D_1 + D_2 - \frac{k_B T}{2\pi\eta r} \tag{41}$$

with $D_1 = D_2 = D_E$ and $\eta = \eta_{int}$ (rather than η_0) to account for the internal viscosity. In Fig. 10, we plot $D(r)$ in units of D_E as a function of r. We see that the increase in D_E due to the intrahelical effects is largely canceled in $D(r)$ from interhelical effects over the range of interest; e.g., $D(a)$ with $a = 16$ Å appearing in the second term of Eq. (30) is equal to 0.5 D_E, which is one-fourth of the value ($2D_E$) expected in the absence of the interhelical HI. For the first term in Eq. (30), the intrahelical effects are a little larger than the interhelical effects so that τ_{UI} (with HI) is about two thirds of τ_{UI} (without HI) (see Table V). Assuming that the value of $l\gamma/\beta$ is not altered by HI, we obtain $k_f \sim 0.093$ ns^{-1} from Eq. (30) (see Table V). This value of k_f is to be compared with that calculated from Eq. (30) with HI neglected (see

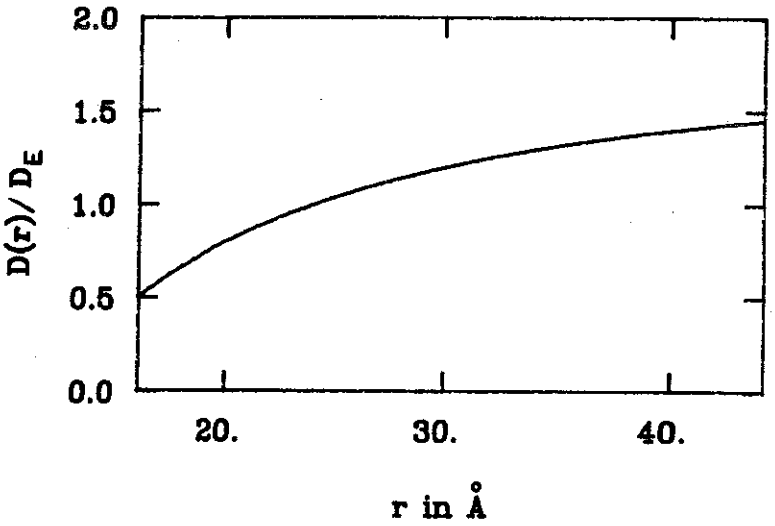


Fig. 10. Variation of the relative diffusion coefficient $D(r)$ with r.

Table V). We see that, when HI is included, the overall folding rate increases by less than a factor of 1.5 in the present model calculation.

It is also of interest to compare these results with those from the earlier calculation based on a simple analytic model.[35] There, it was assumed that each of the helices can be approximated by a sphere whose volume is about the same as that of the helix. The corresponding estimate of the diffusion coefficient for the helix is 30.4 Å^2/ns. This value is approximately the same as that obtained with the effect of HI incorporated (see above). By use of this diffusion coefficient, we obtain $\tau_{UI} = 10.5$ ns from the first term of Eq. (33). This estimate of τ_{UI} is accidentally in good agreement with that calculated from the first term of Eq. (30) with HI effect included (see Table V) because of the compensating errors arising from the neglect of the effect of a realistic intervening chain on $g(r)$ and the interhelical HI effect. Since $\tau_{UI} \gg \tau_{IF}^*$ in the present case, k_f is also correctly determined by the simple model. Whether this fortunate coincidence holds more generally for the microdomain folding problem will have to be determined by additional calculations.

APPENDIX

Statistical Error Limits for the Rate Constants Evaluated from the First-Passage Time Statistics

Consider a transition process, $A \rightarrow B$. If we observe n first-passage times from A to B, $\{t_1, t_2, \ldots, t_n\}$, we may say equivalently that there occurred n transitions during a time interval $T = \sum_{i=1}^{n} t_i$. If the transition is a Poisson process, the probability of the occurrence of m transitions during a time interval T is given by

$$P(T, m) = \frac{(kT)^m}{m!} e^{-kT} \tag{A1}$$

where k is the transition rate ($k\Delta t$ is the probability of a transition during an

infinitesimal time interval Δt and kT is the average number of transitions during a finite time interval T). The lower and upper limits of kT (say, x and x', respectively) may be estimated from the following equations:[36]

$$\sum_{m=n}^{\infty} \frac{x^m}{m!} e^{-x} = \alpha \tag{A2}$$

$$\sum_{m=0}^{n} \frac{(x')^m}{m!} e^{-x'} = \alpha \tag{A3}$$

where 2α is the significance level of the estimation (i.e., the probability that kT lies between x and x' is $1 - 2\alpha$). We can find the values of x and x' for any given values of n and α from a cumulative Poisson distribution table or from a table of percentage points of the χ^2 distribution, since[37]

$$Q(\chi^2|\gamma) = 1 - P(\chi^2|\gamma) = \sum_{m=0}^{c-1} e^{-x} \frac{x^m}{m!} \tag{A4}$$

where $c = \gamma/2$ (γ even) and $x = \chi^2/2$. For $\gamma > 30$, the following asymptotic expression for $P(\chi^2|\gamma)$ is quite accurate:

$$P(\chi^2|\gamma) \approx P(y) \tag{A5}$$

where $y = (\chi^2 - \gamma)/(2\gamma)^{1/2}$ and $P(y)$ is the normal probability function

$$P(y) = (2\pi)^{-1/2} \int_{-\infty}^{y} e^{-t^2/2}\, dt \tag{A6}$$

Finally, the expectation value of k is given by n/T with the lower and upper limits, x/T and x'/T, respectively, for a given significance level 2α.

References

1. Karplus, M. & Weaver, D. L. (1976) *Nature* **260**, 404–406.
2. Karplus, M. & Weaver, D. L. (1979) *Biopolymers* **18**, 1421–1437.
3. Bashford, D., Weaver, D. L. & Karplus, M. (1984) *J. Biomol. Struct. Dynam.* **1**, 1243–1250.
4. Adam, G. & Delbruck, M. (1968) in *Structural Chemistry and Molecular Biology*, Rich, A. & Davidson, N., Eds., Freeman, San Francisco, pp. 198–215.
5. Weaver, D. L. (1979) *Biophys. Chem.* **10**, 245–251.
6. Weaver, D. L. (1982) *Biopolymers* **21**, 1275–1300.
7. Weaver, D. L. (1984) *Biopolymers* **23**, 675–694.
8. Szabo, A., Shulten, K. & Schulten, Z. (1980) *J. Chem. Phys.* **72**, 4350–4357.
9. Zientara, G. P., Nagy, J. A. & Freed, J. H. (1980) *J. Chem. Phys.* **73**, 5092–5106.
10. Zientara, G. P., Nagy, J. A. & Freed, J. H. (1982) *J. Phys. Chem.* **86**, 824–832.
11. McCammon, J. A., Northrup, S. H., Karplus, M. & Levy, R. M. (1980) *Biopolymers* **19**, 2033–2045.
12. Hol, W. G. J., Duijen, P. T. van & Berendsen, H. J. C. (1978) *Nature* **273**, 443–446.
13. Hol, W. G. J., Halie, L. M. & Sander, C. (1981) *Nature* **294**, 532–536.
14. Ryckaert, J. P., Cicotti, G. & Berendsen, H. J. C. (1977) *J. Comput. Phys.* **23**, 327–341.
15. Gunsteren, W. F. van & Karplus, M. (1981) *Nature* **293**, 677–678.
16. Gunsteren, W. F. van (1980) *Mol. Phys.* **40**, 1015–1019.
17. Levitt, M. (1976) *J. Mol. Biol.* **104**, 59–107.

18. Ermak, D. L. & McCammon, J. A. (1978) *J. Chem. Phys.* **69**, 1352–1360.
19. Lee, S. & Karplus, M. to be published.
20. Weiss, G. H. (1967) *Adv. Chem. Phys.* **13**, 1–18.
21. Helfand, E., Wasserman, Z. R. & Weber, T. A. (1980) *Macromolecules* **13**, 526–533.
22. Helfand, E. (1978) *J. Chem. Phys.* **69**, 1010–1018.
23. Koide, A. & Kihara, T. (1974) *Chem. Phys.* **5**, 34–48.
24. Steiger, U. & Fox, R. F. (1982) *J. Math. Phys.* **23**, 296–310.
25. Schulten, K., Schulten, Z. & Szabo, A. (1981) *J. Chem. Phys.* **74**, 4426–4432.
26. Haas, E., Wilchek, M., Katchalski-Katzir, E. & Steinberg, I. Z. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 1807–1811.
27. Peterlin, A. (1972) *Polymer Lett.* **10**, 101–105.
28. Haas, E., Katchalski-Katzir, E. & Steinberg, I. Z. (1978) *Biopolymers* **17**, 11–31.
29. Garcia de la Torre, J. & Bloomfield, V. A. (1981) *Q. Rev. Biophys.* **14**, 81–139.
30. Garcia Bernal, J. M. & Garcia de la Torre, J. (1980) *Biopolymers* **19**, 751–766.
31. Goldstein, R. F. (1985) *J. Chem. Phys.* **83**, 2390–2397.
32. McQuarrie, D. A. (1976) *Statistical Mechanics*, Harper & Row, New York.
33. Lee, S. & Karplus, M. (1984) *J. Chem. Phys.* **81**, 6106–6118.
34. Deutch, J. M. & Felderhof, B. U. (1973) *J. Chem. Phys.* **59**, 1669–1671.
35. Bashford, D. Ph.D. Thesis (Physics), Tufts University.
36. Beyer, W. H. (1971) *Basic Statistical Tables*, The Chemical Rubber Co., Cleveland, OH, pp. 68–69.
37. Abramowitz, M. & Stegun, I. A. (1972) *Handbook of Mathematical Functions*, Dover, New York, pp. 940–941.

Received July 2, 1986
Accepted September 29, 1986