Theory for the rate of contact formation in a polymer chain with local conformational transitions

Huan-Xiang Zhou[a)]
Department of Physics and Institute of Molecular Biophysics, Florida State University, Tallahassee, Florida 32312

(Received 25 September 2002; accepted 29 October 2002)

I derive an expression for the rate of contact formation between two residues in a polymer chain when both residues undergo native to non-native conformational transitions. A contact can only form when the two residues are brought into contact by interresidue diffusion and are in the native conformations at the same time. The entropy of the chain connecting the two residues are accounted for by the potential of mean force for the interresidue distance. Both the equilibrium probabilities of the native conformations and the time scales of the transitions between the native and non-native conformations are important in determining the contact formation rate. For protein residues, transitions between native and non-native conformations occur in the picoseconds time scale. In comparison, the average time for traversing the distance of a few Å by interresidue diffusion is in the nanoseconds time scale. This separation of time scales between residue conformational transitions and interresidue diffusion ensures that the rate of contact formation is much higher than what is expected from the equilibrium probability for both residues to be in the native conformations. © *2003 American Institute of Physics.* [DOI: 10.1063/1.1531588]

I. INTRODUCTION

Contact formation is an elemental step in protein folding. The formation of a nonlocal native contact requires both large-scale motion of the polypeptide chain as well as local conformational rearrangements. In this paper, I present a simple theory for the rate of contact formation which accounts for the large-scale motion of the polypeptide chain by diffusion and local rearrangements as stochastic transitions between native and non-native conformations. The theory may be applied to study the kinetics of β-hairpin[1,2] and coiled–coil[3] formation and have implications for protein folding in general.

The starting point of the present theory is the work of Szabo, Schulten, and Schulten (SSS).[4] SSS calculated the rate of contact formation in a polymer chain as the inverse of the mean first passage time for diffusion toward the contact distance a in the potential of mean force of the end-to-end distance, r. For a Gaussian chain with $\langle r^2\rangle^{1/2} \gg a$, the rate constant was found to be

$$k = 3(6/\pi)^{1/2} D a / \langle r^2 \rangle^{3/2}, \tag{1}$$

where D is the diffusion constant. This result has been used to model contact formation in cytochrome c (Refs. 5 and 6) and in short peptides.[7]

The formation of a native contact between two residues in a protein requires them to have specific backbone and side chain torsional angles, which are reached through local conformational fluctuations. This key fact is recognized in the present theory. For simplicity, the local conformational fluctuations are modeled as stochastic transitions between native and non-native conformations,

$$\text{Native} \underset{\omega_+}{\overset{\omega_-}{\leftrightarrows}} \text{Non-Native}. \tag{2}$$

It is assumed that, when diffusion brings the two residues together, they immediately form the native contact if both of them happen to be in their native conformations. For future reference, I define the relaxation time of the conformational transitions as

$$\tau = (\omega_+ + \omega_-)^{-1} \equiv \omega^{-1}. \tag{3}$$

Langevin and molecular dynamics simulations show that the relaxation time for backbone and side chain conformational fluctuations are in the order of 1–10 ps.[8–10] This time scale is much shorter than the average time to traverse the distance of a few Å by interresidue diffusion with a diffusion constant $\sim 10^{-6}$ cm^2/s, which is a few nanoseconds. The separation in time scales between local conformational transitions and interresidue diffusion has a crucial consequence. The slowing down of contact formation by the small probabilities for the residues to be in their native conformations is much less than what is expected from an equilibrium argument. This outcome is reminiscent of the effect of a stochastic gate on ligand binding rates.[10,11]

In Sec. II, I will derive the general expression for the rate of contact formation in a polymer chain. The results in the limits of slow and fast local conformational transitions are then presented in Sec. III. This is followed by specialization of the theory to a Gaussian chain in Sec. IV. Finally, implications of the present theory for protein folding are discussed in Sec. V.

a)Tel: (850) 644-7052; Fax: (850) 644-0098; Electronic mail: hxzhou@csit.fsu.edu

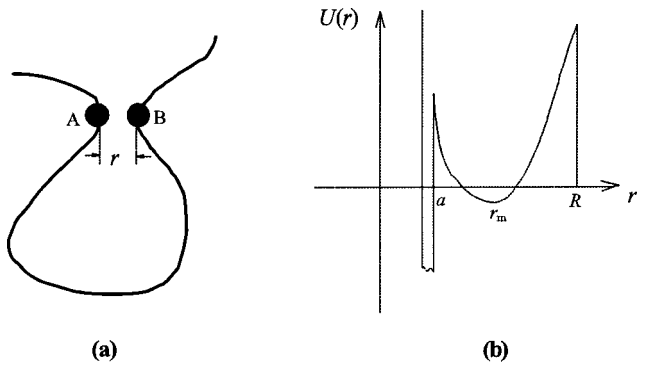

FIG. 1. (a) The contact formation model. (b) The potential of mean force. The deep well inside the contact distance $r=a$ ensures that the contact is formed instantaneously if both residues are in the native conformations.

II. GENERAL THEORY

I consider the contact formation between two residues, A and B, in a polymer chain illustrated in Fig. 1. The interresidue distance, r, will be the reaction coordinate. When the two residues reach the contact distance a, if both of them happen to be in their native conformations, short-range interactions (e.g., hydrophobic and van der Waals interactions and hydrogen bonding) are supposed to provide a deep potential well, such that the residues form a native contact instantaneously.

Outside the contact distance, interresidue motion is modeled as diffusion under the potential of mean force $U(r)$. The potential of mean force is assumed to have a single minimum at a distance r_m between $r=a$ and an outer boundary at $r=R$, which can be infinity. The probability density $p(r,t)$ in r is governed by the Smoluchowski equation,

$$\frac{\partial p(r,t)}{\partial t}=D\frac{\partial}{\partial r}e^{-\beta U(r)}\frac{\partial}{\partial r}e^{\beta U(r)}p(r,t)\equiv\mathcal{L}p(r,t), \tag{4}$$

where $\beta=(k_BT)^{-1}$. Note that a geometric factor r^2 normally appearing in Eq. (4) is absorbed by the potential of mean force. Starting from an equilibrium distribution $e^{-\beta U(r)}/\int_a^R e^{-\beta U(r)}dr$, the total probability $P(t)=\int_a^R p(r,t)dr$ will decrease with time since contact pairs may be absorbed (modeling formation of the native contact). If $P(t)$ decays exponentially, the formation of the native contact can be described as a rate process and the exponent gives the rate constant.

A. Both residues locked in native conformations

Let us first consider the simpler case where the two residues are locked in their native conformations. Then at the contact distance,

$$p(a,t)=0. \tag{5}$$

This is the problem dealt with by SSS. These authors calculated the rate constant as the inverse of the mean first passage time. Let me briefly summarize this approach. The first passage time $\tau(r)$ starting from distance r satisfies the joint equation,

$$De^{\beta U(r)}\frac{d}{dr}e^{-\beta U(r)}\frac{d}{dr}\tau(r)\equiv\mathcal{L}^\dagger\tau(r)=-1. \tag{6}$$

Integrating Eq. (6) with an absorbing boundary condition at $r=a$ [see Eq. (5)] and a reflecting boundary condition at the outer boundary condition at the outer boundary $r=R$, it is found

$$D\tau(r)=\int_a^r e^{\beta U(r)}\int_r^R e^{-\beta U(r')}dr'dr. \tag{7}$$

Averaging over an equilibrium distribution of initial distances and equating the inverse of the mean first passage time as the rate constant leads to

$$k_{\mathrm{SSS}}=D\int_a^R e^{-\beta U(r)}dr\Big/\int_a^R e^{\beta U(r)}\left[\int_r^R e^{-\beta U(r')}dr'\right]^2dr. \tag{8}$$

An alternative is the original approach of Kramers[12] for the rate of diffusive barrier crossing. In this approach, the rate constant is the constant flux from the steady-state probability density with values set to zero at $r=a$ and the equilibrium value at the bottom of the potential well $r=r_m$.[13] The result is[14]

$$k_K=D\Big/\int_a^R e^{-\beta U(r)}dr\int_a^{r_m}e^{\beta U(r)}dr. \tag{9}$$

B. One residue undergoing conformational transitions

Now consider the case where residue B fluctuates between native and non-native conformations. The transitions between the two are assumed to be stochastic processes [see Eq. (2)]. Let the probability densities at r with residue B in the native (n) and non-native (u) conformations be $p_n(r,t)$ and $p_u(r,t)$, respectively. $p_n(r,t)$ and $p_u(r,t)$ will be represented by the vector $\mathbf{p}(r,t)$. Then,

$$\partial\mathbf{p}(r,t)/\partial t=\mathcal{L}\mathbf{p}(r,t)-\tilde{W}\mathbf{p}(r,t), \tag{10a}$$

where

$$\tilde{W}=\begin{bmatrix}\omega_- & -\omega_+\\ -\omega_- & \omega_+\end{bmatrix}. \tag{10b}$$

The inner boundary conditions are

$$p_n(a,t)=0, \tag{11a}$$

$$\left[\frac{\partial}{\partial r}e^{\beta U(r)}p_u(r,t)\right]_{r=a}=0. \tag{11b}$$

That is, the boundary is absorbing for p_n but reflecting for p_u. The outer boundary is reflecting for both $p_n(r,t)$ and $p_u(r,t)$.

The mean first time $\tau_{n,u}(r)$ starting from the native or non-native conformation satisfies

$$\mathcal{L}^\dagger\boldsymbol{\tau}(r)-\tilde{W}^T\boldsymbol{\tau}(r)=-1, \tag{12}$$

where $\tilde{W}^T$ is the transpose of $\tilde{W}$. Let the equilibrium probabilities of the native and non-native conformations be f_n and f_u, which are ω_-/ω and ω_+/ω, respectively. With two new variables,

$$\tau(r) = f_n \tau_n(r) + f_n \tau_u(r), \tag{13a}$$

$$\Delta(r) = \tau_n(r) - \tau_u(r), \tag{13b}$$

Eq. (12) is transformed to

$$\mathcal{L}^\dagger \tau(r) = -1, \tag{14a}$$

$$\mathcal{L}^\dagger \Delta(r) - \omega \Delta(r) = 0. \tag{14b}$$

A function that plays a critical role in the present paper is defined by

$$\mathcal{L}^\dagger q(r;\omega) - \omega q(r;\omega) = 0 \tag{15}$$

with $q(a) = 1$ and a reflecting boundary condition at $r = R$. It is obvious

$$\Delta(r) = \Delta(a) q(r;\omega). \tag{16a}$$

Integrating Eq. (14a), it is found

$$D\tau(r) = C + \int_a^r e^{\beta U(r)} \int_r^R e^{-\beta U(r')} dr' dr. \tag{16b}$$

Using the boundary conditions in Eq. (11), the constant C is determined as

$$C = \frac{\omega_-}{\omega_+} e^{\beta U(a)} \int_a^R e^{-\beta U(r)} dr / [-dq(r;\omega)/dr]_{r=a}. \tag{17}$$

Averaging over an equilibrium distribution of initial distances and equating the inverse of the mean first passage time as the rate constant k_g lead to

$$\frac{1}{k_g} = \frac{1}{k_{\mathrm{SSS}}} + \frac{\omega_-}{\omega_+} \frac{1}{k_K \kappa(\omega)}, \tag{18a}$$

where

$$\kappa(\omega) = e^{-\beta U(a)} \int_a^{r_m} e^{\beta U(r)} dr [-dq(r;\omega)/dr]_{r=a}. \tag{18b}$$

This function is related to the mean first passage time of a particle diffusing in the potential of mean force with an absorbing boundary at $r = a$ and a decay rate ω for $r > a$. If $k_{\mathrm{SSS}} \approx k_K$ are viewed as the "ungated" rate constant, the form of Eq. (18a) is analogous to expressions for rate constants of bimolecular binding affected by conformational transitions[15] or a stochastic gate.[10,11]

C. Both residues undergoing conformational transitions

Generalization to the case where both residues A and B undergo conformational transitions is straightforward. Let the transition rates between native and non-native conformation of residues A and B be ω_{A^-}, ω_{A^+}, ω_{B^-}, and ω_{B^+} [see Eq. (2)]. The expression for the rate constant under conformational transitions of both residues A and B is

$$\frac{1}{k_g} = \frac{1}{k_{\mathrm{SSS}}} + \frac{\omega_{\mathrm{A}^-}}{\omega_{\mathrm{A}^+} k_K \kappa(\omega_{\mathrm{A}})} + \frac{\omega_{\mathrm{B}^-}}{\omega_{\mathrm{B}^+}, k_K \kappa(\omega_{\mathrm{B}})} + \frac{\omega_{\mathrm{A}^-}\omega_{\mathrm{B}^-}}{\omega_{\mathrm{A}^-}\omega_{\mathrm{B}^+} k_K \kappa(\omega_{\mathrm{A}} + \omega_{\mathrm{B}})}, \tag{19}$$

where $\omega_{\mathrm{A}} = \omega_{\mathrm{A}^+} + \omega_{\mathrm{A}^-} = 1/\tau_{\mathrm{A}}$ and $\omega_{\mathrm{B}} = \omega_{\mathrm{B}^+} + \omega_{\mathrm{B}^-} = 1/\tau_{\mathrm{B}}$. The derivation is given in the Appendix.

III. SLOW AND FAST LOCAL CONFORMATIONAL TRANSITIONS

The two limits where the relaxation times τ_{A} and τ_{B} of the local conformational transitions are extremely long or short are of particular interest. When τ_{A} and $\tau_B \to \infty$ (i.e., $\omega = \omega_{\mathrm{A}}$, ω_{B}, or $\omega_{\mathrm{A}} + \omega_{\mathrm{B}} \to 0$), $q(r;\omega)$ can be expanded in powers of ω,

$$q(r;\omega) = q_0(r) + \omega q_1(r) + \cdots. \tag{20}$$

Inserting Eq. (20) into Eq. (15) and solving the resulting equations order by order, it is found $q_0(r) = 1$ and

$$-De^{-\beta U(r)} dq_1(r)/dr = \int_r^R e^{-\beta U(r)} dr. \tag{21}$$

Thus to the lowest order in ω_{A}, ω_{B}, and $\omega_{\mathrm{A}} + \omega_{\mathrm{B}}$,

$$\frac{1}{k_g} = \frac{\omega_{\mathrm{A}^-}}{\omega_{\mathrm{A}^+}\omega_{\mathrm{A}}} + \frac{\omega_{\mathrm{B}^-}}{\omega_{\mathrm{B}^+}\omega_{\mathrm{B}}} + \frac{\omega_{\mathrm{A}^-}\omega_{\mathrm{B}^-}}{\omega_{\mathrm{A}^+}\omega_{\mathrm{B}^+}(\omega_{\mathrm{A}} + \omega_{\mathrm{B}})}. \tag{22}$$

This result has a simple explanation. When τ_{A} and $\tau_{\mathrm{B}} \to \infty$, the fraction of chains initially having one of the residues in the non-native conformation will be stuck in that conformation and will never have a chance to form a native contact. In that limit the mean lifetime is infinite. As τ_{A} and τ_{B} becomes finite, these chains slowly switch to have both residues in the native conformations, thereupon forming a native contact instantaneously (on the time scale of τ_{A} and τ_{B}). In particular, for the case where residue B is locked in the native conformation (i.e., $\omega_{\mathrm{B}^-}/\omega_{\mathrm{B}^+} = 0$), the equilibrium fraction of chains with residue A in the non-native conformation is $\omega_{\mathrm{A}^-}/\omega_{\mathrm{A}}$, and these chains have a mean lifetime of $1/\omega_{\mathrm{A}^+}$ before residue A switches to the native conformation.

When the local conformational transitions are fast (i.e., $\omega \to \infty$), the $q(r;\omega)$ function will decay rapidly from 1 to 0 near $r = a$. Hence $[-dq(r;\omega)/dr]_{r=a} \to \infty$ and $\kappa(\omega) \to \infty$. In this limit, $k_g \to k_{\mathrm{SSS}}$. This result, again analogous to what happens for bimolecular binding affected by conformational transitions[15] or a stochastic gate,[10,11] can be rationalized by the separation in time scales between interresidue diffusion and local conformational transitions. When diffusion brings the two residues together, the slow diffusional steps will leave the two residues in near contact distances for a relatively long time. During this period, fast transitions will have brought the two residues into and out of their native conformations many times. It takes only one event in which both residues are in their native conformations simultaneously in order for the native contact to form.

The separation of times scales between interresidue diffusion and local conformational transitions can be made more concrete. It is easier to work with a new variable $y(r;\omega) = e^{\beta[U(a)-U(r)]/2} q(r;\omega)$. Equation (15) becomes

$$d^2 y(r;\omega)/dr^2 - \{\omega/D + [\beta U'(r)/2]^2 - \beta U^n(r)/2\} y(r;\omega) = 0. \tag{23}$$

Since $\omega \to \infty$ and $q(r;\omega)$ rapidly decays from 1 to 0 near $r=a$, the appropriate solution of Eq. (23) has the form,

$$y(r;\omega)=e^{-(\omega/D)^{1/2}(r-a)}[y_0(r)+(\omega/D)^{-1/2}y_1(r)+\cdots]. \tag{24}$$

Inserting Eq. (24) into Eq. (23) and solving the resulting equations order by order, it is easy to find $y_0(r)=1$. Then,

$$-[dq(r;\omega)/dr]_{r=a}=(\omega/D)^{1/2}-\beta U'(a)/2+\cdots. \tag{25}$$

The parameter,

$$\gamma=\frac{(w/D)^{1/2}}{-\beta U'(a)/2} \tag{26}$$

measures the separation of time scales.

IV. APPLICATION TO THE GAUSSIAN CHAIN

For a Gaussian chain, the potential of mean force for the interresidue distance is given by

$$e^{-\beta U(r)}=r^2e^{-3r^2/2\langle r^2\rangle}. \tag{27}$$

The minimum of the potential is located at

$$r_m=(2\langle r^2\rangle/3)^{1/2} \tag{28}$$

and the outer boundary is at infinity. When $\langle r^2\rangle^{1/2}\sim r_m \gg a$,

$$\int_a^\infty e^{-\beta U(r)}dr=\int_a^\infty r^2e^{-3r^2/2\langle r^2\rangle}dr \approx \int_0^\infty r^2e^{-3r^2/2\langle r^2\rangle}dr=(\pi/6)^{1/2}\langle r^2\rangle^{3/2}/3,$$

and

$$\int_a^{r_m} e^{\beta U(r)}dr=\int_a^{r_m} r^{-2}e^{r^2/r_m^2}dr =\int_{1/r_m}^{1/a} e^{r^2/r_m^2}d(1/r) =(e^{a^2/r_m^2}/a-e/r_m)+(2/r_m^2)\int_a^{r_m} e^{r^2/r_m^2}dr\approx 1/a. \tag{29}$$

Equation (9) for the ''ungated'' rate constant then reduces to Eq. (1), the SSS result.

The mean square of the end-to-end distance for a Gaussian-chain is proportional to the sequence separation n,

$$\langle r^2\rangle=nb_{\text{eff}}^2. \tag{30}$$

The effective bond length b_{eff} for a denatured protein chain depends on the total number N of residues and increases from 6 to 8.5 Å as N increases from 20 to 100.[16] $\langle r^2\rangle^{1/2}=n^{1/2}b_{\text{eff}}=26$ Å for $b_{\text{eff}}=6$ Å and $n=19$ and 67 Å for $b_{\text{eff}}=8.5$ Å and $n=62$. With $D=5\times10^{-7}$ cm^2/s and $a=4$ Å, Eq. (1) predicts $k=4.7\times10^6$ and 2.8×10^5 s^{-1} for the above two values of $\langle r^2\rangle^{1/2}$. These estimated values of k are comparable to the measured rates of 7.2×10^6 s^{-1} for the end-to-end contact formation of a peptide with $n=19$[7b] and $\sim10^5$ s^{-1} for the contact formation between the heme attached to His18 and Met80 in cytochrome c.[6]

The slope of the potential at $r=a$ is given by

$$-\beta U'(a)=(2/a)(1-a^2/r_m^2). \tag{31a}$$

When $r_m \gg a$, $-\beta U'(a)\approx 2/a$ and the time-scale separation parameter defined by Eq. (26) is

$$\gamma(a^2/D\tau)^{1/2}=(\tau_D/\tau)^{1/2}, \tag{31b}$$

where $\tau_D=a^2/D$ is the average time to traverse a distance of a by interresidue diffusion. As noted in the Introduction, τ_D is in the range of nanoseconds, whereas τ is in the range of picoseconds.

Using Eqs. (25), (27), (29), and (31) in Eq. (18b), we find

$$\kappa(\omega)=\gamma+1=(\tau_D/\tau)^{1/2}+1. \tag{32}$$

The rate constant for contact formation is

$$\frac{k}{k_g}=1+\frac{\omega_{A^-}}{\omega_{A^+}[(\tau_D/\tau_A)^{1/2}+1]}+\frac{\omega_{B^-}}{\omega_{B^+}[(\tau_D/\tau_B)^{1/2}+1]} +\frac{\omega_{A^-}\omega_{B^-}}{\omega_{A^+}\omega_{B^+}[(\tau_D/\tau_A+\tau_D/\tau_B)^{1/2}+1]}. \tag{33}$$

When $\omega_{A^-}/\omega_{A^+}$ and $\omega_{B^-}/\omega_{B^+} \gg 1$, the second and third terms on the right-hand side of Eq. (33) are much smaller than the last term. If $\omega_{A^-}/\omega_{A^+}=\omega_{B^-}/\omega_{B^+}=15$, argument based on the equilibrium populations of the native conformations would predict k_g to be smaller than k by 256-fold. In contrast, Eq. (33) predicts a value of k_g that is smaller than k by just sevenfold if $\tau_D/\tau_A=\tau_D/\tau_B=1000$.

V. IMPLICATIONS FOR PROTEIN FOLDING

The formation of a native contact between two residues in protein folding requires the residues to be in their native conformations. The theory developed here makes it clear that both the equilibrium probabilities of the native conformations and the time scales of the transitions between the native and non-native conformations are important in determining the contact formation rate. In particular, the slowing down of contact formation by the small probabilities for the residues to be in their native conformations is much less than what is expected from an equilibrium argument, thanks to the separation of time scales between local conformational transitions and interresidue diffusion.

Both in the present model and in the scenario leading to Levinthal's paradox, the searches for the ''native structure'' by the protein chain are totally random. However, there are crucial differences in chain topology and dynamics. In the Levinthal-paradox scenario, the protein is a set of independent residues, thus the formation of any substructure is independent of chain separation of the residues involved. In contrast, in the model here, the mean time to form a native contact increases with sequence separation. In addition, different time scales are explicitly built in the present model, but there is only one time scale in the Levinthal-paradox scenario. What rescues protein folding from Levinthal's paradox is cooperative interactions between the residues. In the present model such cooperative interactions are mani-

fested by the deep potential well at contact distance (see Fig. 1) when both residues happen to be in their native conformations.

The present theory extends the SSS work by explicitly accounting for local conformational transitions. As such, the SSS theory is perhaps more appropriate for the formation of ''nonspecific'' contacts, such as those between the heme attached to His18 and other ligand residues in cytochrome *c* (Refs. 5 and 6) or between the ends of a short peptide.[7] The present theory, on the other hand, may be applied to study the formation of specific, native contacts, such as those in a β-hairpin[1,2] or a coiled-coil.[3] The rate for forming the end-to-end contact of a 11-residue peptide has been measured to be $\sim 1.5\times 10^7$ s^{-1},[7] which is just what is predicted by Eq. (1) for $D=5\times 10^{-7}$ cm^2/s, $a=4$ Å, and $\langle r^2\rangle$ given by Eq. (30) with $b_{\rm eff}=5.5$ Å (see Ref. 16) and $n=10$. On the other hand, the rate of forming a β-hairpin, as monitored by the quenching of the fluorescence of a Trp residue upon contacting the opposite strand ($n\sim 10$), is only 3.3×10^5 s^{-1}.[2] The 45-fold slow down is easily rationalized if both sides of the hairpin must prearrange into locally native structures before the fluorescence-quenching contacts between the Trp residue and the opposite strand can be formed. According to Eq. (33), a value of $k/k_g=45$ is obtained if $\omega_{\rm A^-}/\omega_{\rm A^+}=\omega_{\rm B^-}/\omega_{\rm B^+}=44$ and $\tau_D/\tau_{\rm A}=\tau_D/\tau_{\rm B}=1000$. There has been a flurry of recent efforts to model β-hairpin formation.[17]

Within the present model, the equilibrium constant is

$$K_g=\frac{k_g}{k_{g^-}}=\frac{f_{nn}\int_{a-\delta a}^{a}e^{-\beta U(r)}dr}{\int_a^R e^{-\beta U(r)}dr}, \tag{34}$$

where k_{g^-} is the rate constant for breaking the native contact, $f_{nn}=\omega_{\rm A^-}\omega_{\rm B^-}/\omega_{\rm A}\omega_{\rm B}$ is the equilibrium probability for both residues to be in the native conformations, and δa is the range of relative distance between the residues in the ''folded'' state. When the potential of mean force is given by the Gaussian-chain model, the contact breaking rate is

$$k_{g^-}=\frac{k_g}{k}\,\frac{Da}{f_{nn}\int_{a-\delta a}^{a}e^{-\beta U(r)}dr}. \tag{35}$$

Equation (35) shows that k_{g^-} is independent of the potential of mean force for $r>a$. In particular, k_g is affected by the sequence separation of the two residues (i.e., loop length) but k_{g^-} is not.

Grantcharova *et al.*[18] investigated the association of structural elements along the folding pathway of the src SH3 domain. Mutagenesis has suggested that the contact between the distal hairpin and the diverging turn (which are connected by the n-src loop) is formed in the folding transition state. Upon inserting 10 glycines in the n-src loop, the folding rate is decreased by fourfold but the unfolding rate does not change at all. These results are consistent with the present model for contact formation.

The approach here based on a simple model complements molecular dynamics simulations of contact formation[19] and other protein folding processes.[17] In particular, the simulations can generate parameters (e.g., the diffusion constant and the transition rates between native and non-native local conformations). The model, on the other hand, may provide a framework for analyzing the simulation results.

The present theory may be viewed as a generalization of the diffusion-collision model introduced by Karplus and Weaver.[20–22] A very important difference is that a potential of mean force is introduced in the present theory. In particular, the entropy of the chain connecting the two contact-forming residues gives rise to such a potential of mean force. In contrast, in the Karplus–Weaver approach, the chain is simply to supply the outer boundary and the residues otherwise undergo free diffusion. Karplus and Weaver[22] essentially considered the situation where only one residue (or microdomain in their language) undergoes conformational transitions, using an expression derived by Bashford.[23] It is more appropriate to treat both residues as independently undergoing conformational transitions, as done in the present theory.

The present theory can be extended in several ways to more realistically model the formation of β-hairpins or coiled-coils and protein folding. For example, a stable native contact may require more than one residue on each side. Formation of one native contact may serve as the nucleus for or otherwise facilitate contact formation of nearby residues. A protein chain may attempt to form more than one native contact,[24] resulting in competition of different pathways. The chain entropy may not be well described by the Gaussian model.[16,25] These effects will be studied in the future.

ACKNOWLEDGMENTS

The author thanks Attila Szabo and Robert L. Baldwin for discussions. This work was supported in part by NIH Grant No. GM58187.

APPENDIX: DERIVATION OF EQ. (19)

Here I derive Eq. (19), the expression for the rate constant of contact formation between two residues when both of them undergo conformational transitions. The conformations of the two residues will be represented by subscripts such as ''*nu*,'' which means residues A and B are in the native and non-native conformations, respectively. The transition matrix for the four combinations of residue conformations (i.e., ''*nn*,'' ''*un*,'' ''*nu*,'' and ''*uu*'') now reads

$$\tilde{W}=\begin{bmatrix} \omega_{\rm A^-}+\omega_{\rm B^-} & -\omega_{\rm A^+} & -\omega_{\rm B^+} & 0 \\ -\omega_{\rm A^-} & \omega_{\rm A^+}+\omega_{\rm B^-} & 0 & -\omega_{\rm B^+} \\ -\omega_{\rm B^-} & 0 & \omega_{\rm A^-}+\omega_{\rm B^+} & -\omega_{\rm A^+} \\ 0 & -\omega_{\rm B^-} & -\omega_{\rm A^-} & \omega_{\rm A^+}+\omega_{\rm B^+} \end{bmatrix}. \tag{A1}$$

After introducing four new variables,

$$\tau(r)=f_{nn}\tau_{nn}(r)+f_{un}\tau_{dn}(r)+f_{nu}\tau_{nu}(r)+f_{uu}\tau_{uu}(r), \tag{A2a}$$

$$\Delta(r)=\tau_{nn}(r)-\tau_{un}(r)-\tau_{nu}(r)+\tau_{uu}(r), \tag{A2b}$$

$$\tau_A(r) = f_{Bn}[\tau_{nn}(r) - \tau_{un}(r)] + f_{Bu}[\tau_{nu}(r) - \tau_{uu}(r)], \tag{A2c}$$

$$\tau_B(r) = f_{An}[\tau_{nn}(r) - \tau_{nu}(r)] + f_{Au}[\tau_{un}(r) - \tau_{uu}(r)], \tag{A2d}$$

it can be shown

$$\mathcal{L}^\dagger \tau(r) = -1, \tag{A3a}$$

$$\mathcal{L}^\dagger \Delta(r) - (\omega_A + \omega_B)\Delta(r) = 0, \tag{A3b}$$

$$\mathcal{L}^\dagger \tau_A(r) - \omega_A \tau_A(r) = 0, \tag{A3c}$$

$$\mathcal{L}^\dagger \tau_B(r) - \omega_B \tau_B(r) = 0. \tag{A3d}$$

The solutions for Eq. (A3) are

$$D\tau(r) = C + \int_a^r e^{\beta U(r)} \int_r^R e^{-\beta U(r')} dr' dr, \tag{A4a}$$

$$\Delta(r) = \Delta(a) q(r; \omega_A + \omega_B), \tag{A4b}$$

$$\tau_A(r) = \tau_A(a) q(r; \omega_A), \tag{A4c}$$

$$\tau_B(r) = \tau_B(a) q(r; \omega_B). \tag{A4d}$$

The four constants C, $\Delta(a)$, $\tau_A(a)$, and $\tau_B(a)$ can be determined by the absorbing condition for $t_{nn}(r)$ and reflecting conditions for $\tau_{un}(r)$, $\tau_{nu}(r)$, and $\tau_{uu}(r)$ at $r=a$. In particular,

$$C = e^{\beta U(a)} \int_a^R e^{-\beta U(a)} dr \left[\frac{\omega_{A^-}}{-\omega_{A^+} q'(a; \omega_A)} + \frac{\omega_{B^-}}{-\omega_{B^+} q'(a; \omega_B)} + \frac{\omega_{A^-}\omega_{B^-}}{-\omega_{A^+}\omega_{B^+} q'(a; \omega_A + \omega_B)} \right]. \tag{A5}$$

Averaging Eq. (A4a) over an equilibrium distribution of initial distances leads to Eq. (19).

[1] V. Munoz, P. A. Thompson, J. Hofrichter, and W. A. Eaton, Nature (London) **390**, 196 (1997).
[2] G. S. Jas, W. A. Eaton, and J. Hofrichter, J. Phys. Chem. B **105**, 261 (2001).
[3] L. B. Moran, J. P. Schneider, A. Kentsis, G. A. Reddy, and T. R. Sosnick, Proc. Natl. Acad. Sci. U.S.A. **96**, 10699 (1999).
[4] A. Sazbo, K. Schulten, and Z. Schulten, J. Chem. Phys. **72**, 4350 (1980).
[5] C. M. Jones, E. R. Henry, Y. Hu, C. Chan, S. D. Luck, A. Bhuyan, H. Roder, J. Hofrichter, and W. A. Eaton, Proc. Natl. Acad. Sci. U.S.A. **90**, 11860 (1993); S. J. Hagen, J. Hofrichter, A. Szabo, and W. A. Eaton, *ibid.* **93**, 11615 (1996); S. J. Hagen, J. Hofrichter, and W. A. Eaton, J. Phys. Chem. B **101**, 2352 (1997).
[6] S. J. Hagen, C. W. Carswell, and E. M. Sjolander, J. Mol. Biol. **305**, 1161 (2001).
[7] (a) O. Bieri, J. Wirz, B. Hellrung, M. Schutkowski, M. Drewello, and T. Kiefhaber, Proc. Natl. Acad. Sci. U.S.A. **96**, 9597 (1999); (b) L. J. Lapidus, W. A. Eaton, and J. Hofrichter, *ibid.* **97**, 7220 (2000).
[8] (a) V. Dagget, P. A. Kollman, and I. D. Kuntz, Biopolymers **31**, 1115 (1991); (b) V. Dagget and M. Levitt, J. Mol. Biol. **223**, 1121 (1992); (c) K. Yapa, D. L. Weaver, and M. Karplus, Proteins **12**, 237 (1992); (d) W. Schneller and D. L. Weaver, Biopolymers **33**, 1519 (1993).
[9] S. T. Wlodek, T. W. Clark, L. R. Scott, and J. A. McCammon, J. Am. Chem. Soc. **119**, 9513 (1997).
[10] H.-X. Zhou, S. T. Wlodek, and J. A. McCammon, Proc. Natl. Acad. Sci. U.S.A. **95**, 9280 (1998).
[11] H.-X. Zhou, J. Chem. Phys. **108**, 8146 (1998).
[12] H. A. Kramers, Physica (Utrecht) **7**, 284 (1940).
[13] H.-X. Zhou, Biopolymers **59**, 427 (2001).
[14] H.-X. Zhou, J. Phys. Chem. B **106**, 2393 (2002).
[15] H.-X. Zhou and A. Szabo, Biophys. J. **71**, 2440 (1996).
[16] H.-X. Zhou, J. Phys. Chem. B **106**, 5769 (2002).
[17] (a) V. Munoz, E. R. Henry, J. Hofrichter, and W. A. Eaton, Proc. Natl. Acad. Sci. U.S.A. **95**, 5872 (1998); (b) V. S. Pande and D. S. Rokhsar, *ibid.* **96**, 9062 (1999); (c) A. R. Dinner, T. Lazaridis, and M. Karplus, *ibid.* **96**, 9068 (1999); D. K. Klimov and D. Thirumalai, *ibid.* **97**, 2544 (2000); C. Guo, H. Levine, and D. A. Kessler, *ibid.* **97**, 10775 (2000); R. Zhou, B. J. Berne, and R. Germain, *ibid.* **98**, 14931 (2001).
[18] V. P. Grantcharova, D. S. Riddle, and D. Baker, Proc. Natl. Acad. Sci. U.S.A. **97**, 7084 (2000).
[19] I.-C. Yeh and G. Hummer, J. Am. Chem. Soc. **124**, 6563 (2002).
[20] M. Karplus and D. L. Weaver, Nature (London) **260**, 404 (1976).
[21] M. Karplus and D. L. Weaver, Biopolymers **18**, 1421 (1979).
[22] M. Karplus and D. L. Weaver, Protein Sci. **3**, 650 (1994).
[23] D. Bashford, J. Chem. Phys. **85**, 6999 (1986).
[24] D. E. Makarov and H. Metiu, J. Chem. Phys. **116**, 5205 (2002); D. E. Makarov, C. A. Keller, K. W. Plaxco, and H. Metiu, Proc. Natl. Acad. Sci. U.S.A. **99**, 3535 (2002).
[25] H.-X. Zhou, J. Phys. Chem. B **105**, 6763 (2001).