Kinetics of Loop Formation in Polymer Chains†

Ngo Minh Toan,¶,‡ Greg Morrison,¶,£ Changbong Hyeon,§ and D. Thirumalai*,¶,#

Biophysics Program, Institute for Physical Science and Technology, University of Maryland at College Park, College Park, Maryland 20742, Department of Physics, University of Maryland at College Park, College Park, Maryland 20742, Center for Theoretical Biological Physics, University of California at San Diego, La Jolla, California 92093, and Department of Chemistry and Biochemistry, University of Maryland at College Park, College Park, Maryland 20742

Received: August 13, 2007; In Final Form: November 8, 2007

We investigate the kinetics of loop formation in ideal flexible polymer chains (the Rouse model), and polymers in good and poor solvents. We show for the Rouse model, using a modification of the theory of Szabo, Schulten, and Schulten, that the time scale for cyclization is $\tau_c \sim \tau_0 N^2$ (where τ_0 is a microscopic time scale and N is the number of monomers), provided the coupling between the relaxation dynamics of the end-to-end vector and the looping dynamics is taken into account. The resulting analytic expression fits the simulation results accurately when a, the capture radius for contact formation, exceeds b, the average distance between two connected beads. Simulations also show that when $a < b$, $\tau_c \sim N^{\alpha_\tau}$, where $1.5 < \alpha_\tau \leq 2$ in the range $7 < N < 200$ used in the simulations. By using a diffusion coefficient that is dependent on the length scales a and b (with $a < b$), which captures the two-stage mechanism by which looping occurs when $a < b$, we obtain an analytic expression for τ_c that fits the simulation results well. The kinetics of contact formation between the ends of the chain are profoundly effected when interactions between monomers are taken into account. Remarkably, for $N < 100$, the values of τ_c decrease by more than 2 orders of magnitude when the solvent quality changes from good to poor. Fits of the simulation data for τ_c to a power law in N ($\tau_c \sim N^{\alpha_\tau}$) show that α_τ varies from about 2.4 in a good solvent to about 1.0 in poor solvents. The effective exponent α_τ decreases as the strength of the attractive monomer−monomer interactions increases. Loop formation in poor solvents, in which the polymer adopts dense, compact globular conformations, occurs by a reptation-like mechanism of the ends of the chain. The time for contact formation between beads that are interior to the chain in good solvents changes nonmonotonically as the loop length varies. In contrast, the variation in interior loop closure time is monotonic in poor solvents. The implications of our results for contact formation in polypeptide chains, RNA, and single-stranded DNA are briefly outlined.

1. Introduction

Contact formation (cyclization) between the ends of a long polymer has been intensely studied both experimentally[1,2] and theoretically.[3−9] More recently, the kinetics of loop formation has become increasingly important, largely because of its relevance to DNA looping[10,11] as well as protein[12−19] and RNA folding.[20] The ease of cyclization in DNA, which is a measure of its intrinsic flexibility,[11,21] is important in gene expression and interactions of DNA with proteins and RNA. In addition, the formation of contacts between residues (nucleotides) near the loop[8] may be the key nucleating event in protein (RNA) folding. For these reasons, a number of experiments have probed the dependence of the rates of cyclization in proteins[12,13,22] and RNA[23,24] as a function of loop length. The experimental reports, especially on the rates of loop formation in polypeptides and proteins, have prompted a number of theoretical studies[7,25,26] that build on the pioneering treatments by Wilemski and Fixman[3] (WF) and Szabo, Schulten, and Schulten[4] (SSS). The WF formalism determines the loop closure time τ_c by solving the diffusion equation in the presence of a sink term. The sink function accounts for the possibility that contact between the ends of a polymer chain occurs whenever they are in proximity. The time for forming a loop is related to a suitable time integral of the sink−sink correlation function.

In an important paper, SSS developed a much simpler theory to describe the dependence of the rate of end-to-end contact formation in an ideal chain on the polymer length N. The SSS approximation[4] describes the kinetics of contact formation between the ends of the chain as a diffusion process in an effective potential that is derived from the probability distribution $P(\mathbf{R}_{ee})$ of finding the chain ends with the end-to-end distance $\mathbf{R}_{ee}$. More recently, such an approach has been adopted to obtain the rates of folding of proteins from a free-energy surface expressed in terms of an appropriately chosen reaction coordinate.[27] The validity of using the dynamics in a potential of mean force, $F(\mathbf{R}_{ee}) \sim -k_B T \log[P(\mathbf{R}_{ee})]$, to obtain τ_c hinges on local equilibrium being satisfied, that is, that all processes except the one of interest must occur rapidly. In the case of cyclization kinetics in simple systems (Rouse model or self-avoiding polymer chains), the local equilibrium approximation depends minimally on the cyclization time τ_c and the internal

† Part of the "Attila Szabo Festschrift".

* To whom correspondence should be addressed.

¶ Biophysics Program, Institute for Physical Science and Technology, University of Maryland at College Park.

‡ Present address: Institute of Physics and Electronics, 10-Dao Tan, Hanoi, Vietnam (on leave).

£ Department of Physics, University of Maryland at College Park.

§ University of California at San Diego.

Department of Chemistry and Biochemistry, University of Maryland at College Park.

10.1021/jp076510y CCC: $40.75
Published on Web 02/13/2008

chain relaxation time τ_R. In the limit $\tau_c/\tau_R \gg 1$, one can envision the motions of the ends as occurring in the effective free energy $F(\mathbf{R}_{ee})$ because the polymer effectively explores the available volume before the ends meet. By solving the diffusion equation for an ideal chain for which $F(\mathbf{R}_{ee}) \sim 3k_BT\mathbf{R}_{ee}^2/2\bar{R}_{ee}^2$, with $\bar{R}_{ee} \sim b\sqrt{N}$, where b is the monomer size, subject to absorbing boundary conditions, SSS showed that the mean first passage time for contact formation ($\sim\tau_c$) is $\tau_{SSS} \sim \tau_0 N^{3/2}$, where τ_0 is a microscopic time constant (see eq 7).

The simplicity of the SSS result, which reduces contact formation kinetics to merely computing $P(\mathbf{R}_{ee})$, has resulted in its widespread use to fit experimental data on polypeptide chains.[12,13,22] The dependence of τ_c on N using the SSS theory differs from the WF predictions. In addition, simulations also show that τ_c deviates from the SSS prediction.[28−31] The slower dependence of τ_{SSS} on N can be traced to the failure of the assumption that all internal chain motions occur faster than the process of interest. The interplay between τ_c and τ_R, which determines the validity of the local equilibrium condition, can be expressed in terms of well-known exponents that characterize equilibration and relaxation properties of the polymer chain. Comparison of the conformational space explored by the chain ends and the available volume prior to cyclization[32] allows us to express the validity of the local equilibrium in terms of $\theta = (d + g)/z$, where d is the spatial dimension, g is the des Cloizeaux correlation hole exponent that accounts for the behavior of $P(\mathbf{R}_{ee})$ for small $\mathbf{R}_{ee}$, that is, $P(\mathbf{R}_{ee}) \sim \mathbf{R}_{ee}^g$, and z is the dynamical scaling exponent ($\tau_R \sim \bar{R}_{ee}^z$). Additional discussions along these lines are given in Appendix A. The SSS assumption is only valid provided that $\theta > 1$.[33] For the Rouse chain in the freely draining limit, ($\nu = 1/2$, $g = 0$, $d = 3$, $z = 4$) gives $\theta < 1$, and hence, τ_c will show deviation from the SSS predictions for all N.

The purpose of this paper is twofold. (i) The theory based on the WF formalism and simulations show the closure time $\tau_{WF} \sim \langle \mathbf{R}_{ee}^2 \rangle / D_c \sim N^{1+2\nu}$ ($\nu \approx 3/5$ for self-avoiding walk and $\nu = 1/2$ for the Rouse chain), where D_c is a diffusion constant. We show that the WF result for Rouse chains, τ_{WF}, can be obtained within the SSS framework provided an effective diffusion constant that accounts for the relaxation dynamics of the ends of the chains is used instead of the monomer diffusion coefficient D_0. Thus, the simplicity of the SSS approach can be preserved while recovering the expected scaling result[3,5] for the dependence of τ_c on N. (ii) The use of the Rouse model may be appropriate for polymers or polypeptide chains near Θ conditions. In both good and poor solvents, interactions between monomers determine the statics and dynamics of the polymer chains. The chain will swell in good solvents ($\nu \approx 3/5$), whereas in poor solvents, polymers and polypeptide chains adopt compact globular conformations. In these situations, interactions between the monomers or the amino acid residues affect τ_c. The monomer−monomer interaction energy scale, ϵ_{LJ}, leading to the chain adopting a swollen or globular conformation, influences both ν and the chain relaxation dynamics and hence affects τ_c. Because analytic theory in this situation is difficult, we provide simulation results for τ_c as a function of ϵ_{LJ} and for $10 < N \leq 100$.

2. Derivation of τ_{WF} for the Rouse Model Using the SSS Approximation

The Rouse chain consists of N beads, with two successive beads connected by a harmonic potential that keeps them at an average separation b (the Kuhn length). Contact formation between the chain ends can occur only if fluctuations result in monomers 1 and N being within a capture radius a. In other words, the space explored by the chain ends must overlap within the contact volume $\sim a^3$. There are three relevant time scales that affect loop closure dynamics, namely: $\tau_0 \approx b^2/D_0$, the fluctuation time scale of a single monomer, τ_{ee}, the relaxation time associated with the fluctuations of the end-to-end distance, and τ_R, the relaxation time of the entire chain. Clearly, $\tau_{ee} < \tau_c \sim \tau_R$. Because loop formation can occur only if the ends can approach each other, processes that occur on time scale τ_{ee} must be coupled to looping dynamics. We obtain the scaling of τ_c with N, found using the WF approximation, from the SSS formalism using a diffusion constant evaluated on the time scale τ_{ee}.

2.1. Fluctuations in $\mathbf{R}_{ee}$. The Langevin equation for a Gaussian chain is[34]

$$\gamma \frac{\partial \mathbf{r}(s,t)}{\partial t} = -\frac{\delta H_0[\mathbf{r}(s,t)]}{\delta \mathbf{r}(s,t)} + \vec{\eta}(s,t) \tag{1}$$

where $\vec{\eta}(s,t)$ is a white noise force with $\langle \vec{\eta}(s,t) \rangle = 0$ and $\langle \vec{\eta}(s,t) \cdot \vec{\eta}(s',t') \rangle = 6\gamma k_B T \delta(t - t')\delta(s - s')$; γ is the friction coefficient, and $D_0 = k_BT/\gamma$ is the microscopic diffusion coefficient. By writing

$$\mathbf{r}(s,t) = \mathbf{r}_0 + 2\sum_{n=1}^{N-1} \mathbf{r}_n(t) \cos(n\pi s/N)$$

the Gaussian Hamiltonian H_0 becomes

$$H_0 = \frac{3}{2b^2}\int_0^N \mathrm{d}s \left(\frac{\partial \mathbf{r}(s,t)}{\partial s}\right)^2 = \frac{3}{2Nb^2}\sum_n n^2\pi^2\mathbf{r}_n^2(t) \tag{2}$$

The equation of motion for each mode

$$\dot{\mathbf{r}}_n(t) = -\frac{3n^2\pi^2 D_0}{N^2 b^2}\mathbf{r}_n(t) + \vec{\eta}_n(t) \tag{3}$$

can be solved independently. The solutions naturally reveal the time scale for global motions of the chain, $\tau_R = N^2b^2/3D_0\pi^2 \sim N^2b^2/D_0$. We note that τ_R is much larger than the relevant time scale for internal motions of the monomers, $\tau_1 \approx b^2/D_0$ for large N. Equation 3 can be solved directly, and the fluctuations in the end-to-end distance $\mathbf{R}_{ee}$ are given by

$$\langle \delta\mathbf{R}_{ee}^2(t) \rangle = 16Nb^2 \sum_{n\ \mathrm{odd}} \frac{N^2}{n^4\pi^4} \sin^2\left(\frac{n\pi}{N}\right)(1 - \mathrm{e}^{-n^2t/\tau_R}) \tag{4}$$

with $\langle \delta\mathbf{R}_{ee}^2(t) \rangle \equiv \langle [\mathbf{R}_{ee}(t) - \mathbf{R}_{ee}(0)]^2 \rangle$. The details of the calculation leading to eq 4 are given in Appendix B. If we define an effective diffusion constant using

$$D(t) = \frac{\langle \delta\mathbf{R}_{ee}^2(t) \rangle}{6t} \tag{5}$$

then $D(0) = 2D_0$, as is expected for the short time limit.[4,30] On time scales on the order of τ_R, we find $D(\tau_R) \sim D_0/N$, which is identical to the diffusion constant for the center of mass of the chain.[34] This is the expected result for the diffusion constant for global chain motion.

2.2. The Effective Diffusion Constant. The theory of Szabo, Schulten, and Schulten[4] (SSS) determines the loop closure time by replacing the difficult polymer problem, having many degrees of freedom, with a single particle diffusing in a potential of

mean force. With this approximation, τ_c, which can be related to the probability that the contact is not formed (see Appendix C for more details), becomes

$$\tau_c = \frac{1}{\mathcal{N}}\int_a^{Nb} \mathrm{d}r\, \frac{1}{D(r)P(r)} \left(\int_r^{Nb} \mathrm{d}r' P(r')\right)^2 + \frac{1}{\kappa \mathcal{N} P(a)} \quad (6)$$

where loop closure occurs when $|\mathbf{R}_{ee}| = a$, the closure (or capture) radius, with rate κ, $P(r)$ is the equilibrium end-to-end distribution of the chain, and

$$\mathcal{N} = \int_a^{Nb} \mathrm{d}r\, P(r)$$

In this paper, we will consider only a chemically irreversible process, with the binding rate constant of $\kappa \to \infty$. In the case of the noninteracting Gaussian chain, $P(r) \sim r^2\exp(-3r^2/2Nb^2)$. If $D(r) \sim D_0$ is a constant, it is simple to show[4] that, for large N, the loop closure time is

$$\tau_{SSS} \approx \frac{1}{3}\sqrt{\frac{\pi}{6}}\frac{N^{3/2}b^3}{D_0 a} \quad (7)$$

The scaling of τ_{SSS} with N given in eq 7 disagrees with other theories[3,7] and numerous simulations[28−31] that predict $\tau_c \sim N^2$ for $Nb^2 \gg a^2$ and $a \geq b$. It has been noted[25,33] that the SSS theory may be a lower bound on the loop closure time for a freely draining Gaussian chain, and that an effective diffusion coefficient that is smaller than D_0 is required to fit the simulated[25] and experimental[35] data using τ_{SSS}. Physically, the use of a smaller diffusion constant is needed because contact formation requires fluctuations that bring $|\mathbf{R}_{ee}|$ within the capture radius a, a mechanism in which τ_{ee} plays a crucial role.

As noted by Doi,[5] the relevant time scale for loop closure is not simply the global relaxation time. The fluctuations in $\mathbf{R}_{ee}$ are given not only by the longest relaxation time but also from important contributions that arise from higher modes. This gives rise to the differences between the Harmonic Spring and Rouse models.[5,29] In the Harmonic Spring model, the chain is replaced with only one spring which connects the two ends of the chain. The spring constant is chosen to reproduce the end-to-end distribution function. The higher-order modes give rise to excess fluctuations in $\langle \mathbf{R}_{ee}^2 \rangle$ on a scale of $\sim 0.4\sqrt{N}b = R'$, and their inclusion is necessary to fully capture the physics of loop closure. In the approximation of a particle diffusing in an effective potential (as in the SSS theory), this time scale is simple to determine. If we consider only the x component of $\mathbf{R}_{ee}$, we can treat it as a particle diffusing in a potential $U_{eff}(R_x) = 3R_x^2/2Nb^2 - O(1)$, with diffusion constant $D = 2D_0$. In this case, we find

$$\langle \delta R_x^2(t) \rangle = \frac{2}{3} Nb^2 (1 - \mathrm{e}^{-t/\tau_{ee}}) \quad (8)$$

and $\langle \mathbf{R}_{ee}^2(t) \rangle = 3\langle \delta R_x^2(t) \rangle$, giving the natural end-to-end relaxation time $\tau_{ee} = Nb^2/6D_0$. Because we have evaluated τ_{ee} using diffusion in an effective potential, the dependence of τ_{ee} on N should be viewed as a mean field approximation.

We can determine the effective diffusion constant on the time scale τ_{ee}, which includes the relaxation of $\mathbf{R}_{ee}(t)$ at the mean field level. We define the effective diffusion constant as

$$D_{ee} = \lim_{t \sim \tau_{ee}} \frac{\langle \delta \mathbf{R}_{ee}^2(t) \rangle}{6t} \quad (9)$$

with $\langle \delta \mathbf{R}_{ee}^2(t) \rangle$ in eq 4, which includes all of the modes of the chain and not simply the lowest one. Noting that $\tau_{ee}/\tau_R \sim N^{-1} \ll 1$ for large N, we can convert the sum in eq 4 into an integral

$$\langle \delta \mathbf{R}_{ee}^2 \rangle \approx \frac{2\sqrt{2}}{\pi} N^{3/2} b^2 \int_0^\infty \mathrm{d}x\, \frac{\sin^2(bx/\sqrt{3D_0 t})}{x^4}(1 - \mathrm{e}^{-x^2}) \quad (10)$$

$$\approx 8b\sqrt{\frac{3D_0 t}{\pi}} \quad (11)$$

In particular, for $t \approx \tau_{ee}/2 = Nb^2/12D_0$

$$D_{ee} \approx \frac{8D_0}{\sqrt{N\pi}} - \frac{16D_0}{3N} + O(N^{-3/2}) \quad (12)$$

We expect these coefficients to be accurate to a constant on the order of unity. The effective diffusion constant D_{ee} takes the higher-order modes of the chain into account and should capture the essential physics of the loop closure. In other words, on the time scale τ_{ee}, resulting in $D_{ee} \sim N^{-(1/2)}$, the monomers at the chain ends are within a volume of $\sim a^3$, so that contact formation is possible.

Substituting D_{ee} into eq 7 gives

$$\tau_c \approx \frac{N^2 b^3 \pi}{24\sqrt{6}\, D_0 a} \sim \tau_{WF} \quad (13)$$

in the limit of large N. Thus, within the SSS approximation, the N^2 dependence of τ_c may be obtained, provided the effective diffusion constant D_{ee} is used. The importance of using a diffusion constant that takes relaxation dynamics of $\mathbf{R}_{ee}$ into account has also been stressed by Portman.[25] The closure time in eq 13 depends on the capture radius as a^{-1}, which disagrees with the a-independent prediction of Doi.[5] In addition, eq 13 does not account for the possibility of $\tau_c \sim N^{\alpha_\tau}$, with $1.5 < \alpha_\tau < 2$, as observed with simulations by Pastor et al.[28] when the capture radius is $a < b$. Both of these discrepancies are discussed in the next section by using insights garnered from simulations.

2.3. Simulations of Loop Closure Time for Freely Jointed Chains. In order to measure $\mathbf{R}_{ee}(t)$ and τ_c for a noninteracting freely jointed chain, we have performed extensive Brownian dynamics simulations. We model the connectivity of the chain using the Hamiltonian

$$\beta H = \frac{k_s}{2}\sum_{i=1}^{N}\left(1 - \frac{|\mathbf{r}_{i+1} - \mathbf{r}_i|}{b_0}\right)^2 \quad (14)$$

with $b_0 = 0.38$ nm and a spring constant of $k_s = 100$. We note that $\langle (\mathbf{r}_{i+1} - \mathbf{r}_i)^2 \rangle^{1/2} \approx 0.39$ nm for this Hamiltonian, which we take as the Kuhn length b when fitting the data. For large N, the differences between the FJC and Rouse models are not relevant, and hence, the scaling of τ_c with N for these two models should be identical. The microscopic diffusion coefficient was taken as $D_0 = 0.77$ nm^2/ns. The equations of motion in the overdamped limit were integrated using the Brownian dynamics algorithm,[36] with a time step of $\Delta t = 10^{-4}$ ns. The end-to-end distribution $P(r)$ is easily computed for the model in eq 14, giving the expression for large k_s

$$P(r) = 2r\int_0^\infty \mathrm{d}q\, q \sin(qr)\left(\frac{e^{-b_0^2 q^2/k_s^2}}{b_0 q(1 + k_s)} \times [b_0 q\cos(b_0 q) + k_s \sin(b_0 q)]\right)^{N-1} \quad (15)$$

which must be numerically integrated.

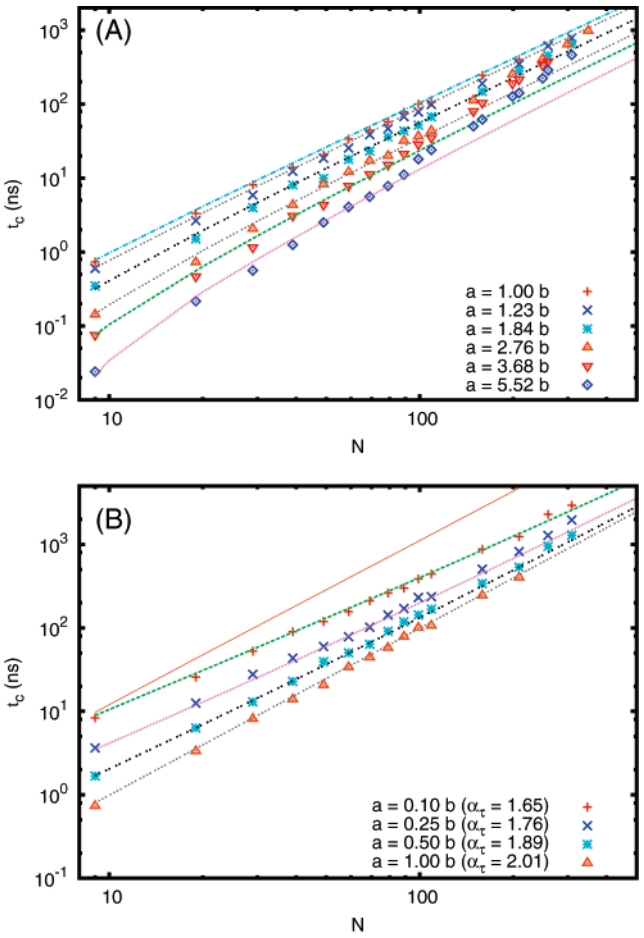

Figure 1. Dependence of τ_c on N for various values of a. The symbols correspond to different values of the capture radius. (A) The values of a/b are 1.00 (+), 1.23 (×), 1.84 (*), 2.76 (Δ), 3.68 (∇), and 5.52 (◊). The lines are obtained using eq 6 with $\kappa \rightarrow \infty$. The diffusion constant in eq 6 is obtained using $D = \langle \delta \mathbf{R}_{ee}^2(\tau_{ee}/2)/3\tau_{ee} \rangle$, with $\langle \delta \mathbf{R}_{ee}^2(t) \rangle$ given in eq 10. (B) The values of a/b are 0.10 (+), 0.25 (×), 0.50 (*), and 1.00 (Δ). The lines are the theoretical predictions using eq 17. The poor fit using eq 13 with $a = 0.1b$ (solid line) shows that the two-stage mechanism has to be included to obtain accurate values of τ_c. The effective exponent α_τ, obtained by fitting $\tau_c \sim N^{\alpha_\tau}$, is shown in parentheses.

In our simulations, we computed the mean first passage time directly. We generated the initial conditions by Monte Carlo equilibration. Starting from each equilibrated initial configuration, the equations of motion were integrated until $|\mathbf{R}_{ee}| \leq a$ for the first time, with the first passage time computed for multiple values of N and a. The loop closure time τ_c was identified with the mean first passage time, obtained by averaging over 400 independent trajectories. For comparison with the analytic theory, we calculated the modified SSS first passage time, with $P(r)$ given in eq 15 and D_{ee} given in eq 12. The results are shown in Figure 1. We find that the behavior of τ_c depends strongly on the ratio a/b.

$a \geq b$: For $N \lesssim 100$ and $a \geq b$, we find that the modified SSS theory using the effective diffusion constant D_{ee} in eq 12 gives an excellent fit to the data as a function of both N and a (Figure 1A). Thus, modeling the loop closure process as a one-dimensional diffusive process in a potential of mean force is appropriate as long as a diffusion coefficient that takes the dynamics of the chain ends into account is used.

For $N \gtrsim 100$ and $a \geq b$, we notice significant deviations in the data from the theoretical curves. The data points appear to converge as a is varied for large N, suggesting the emergence

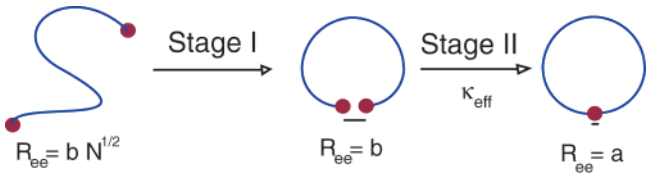

Figure 2. Sketch of the two-stage mechanism for loop closure for Rouse chains when $a < b$. Although unphysical, this case is of theoretical interest. In the first stage, fluctuations in $\mathbf{R}_{ee}$ result in the ends approaching $|\mathbf{R}_{ee}| = b$. The search of the monomers within a volume of b^3 ($>a^3$), which is rate limiting, leads to a contact in the second stage.

of Doi's[5] predicted scaling of $\tau_c \sim N^2a^0$. This departure from the predictions of eq 13 suggests that the one-dimensional mean field approximation, which gives rise to the a dependence of τ_c, breaks down. Even our modified theory, which attempts to include fluctuations in $\mathbf{R}_{ee}$ on a mean field level leading to D_{ee}, cannot accurately represent the polymer as a diffusive process with a single degree of freedom for large N. In this regime, the many degrees of freedom of the polymer must be explicitly taken into account, making the WF theory[3] more appropriate.

$a < b$: The condition $a < b$ is nonphysical for a freely jointed chain with excluded volume and certainly not relevant for realistic flexible chains, in which an excluded volume interaction between monomers would prevent the approach of the chain ends to distances less than b. (Note that for wormlike chains, with the statistical segment $l_p > b$, the equivalent closure condition $a < l_p$ is physically realistic. The effect of chain stiffness, which has been treated elsewhere,[33] is beyond the scope of this article.) In this case (Figure 1B), we find $\tau_c \sim N^{\alpha_\tau}$, with $1.5 < \alpha_\tau < 2$, in agreement with the simulation results of Pastor et al.[28] In deriving D_{ee}, we assumed, as did Doi,[5] that the relaxation of the end-to-end vector is rate limiting. Once $|\mathbf{R}_{ee}| \sim R' \approx 0.4\sqrt{N}b$, we expect the faster internal motions of the chain will search the conformational space rapidly, so that τ_c is dominated by the slower, global motions of the chain (i.e., it is diffusion limited). This assumption breaks down if $a \ll b$ because the endpoints must search longer for each other using the rapid internal motions on a time scale of b^2/D_0. In the limit of small a, the memory of the relaxation of the ends of the chain is completely lost. Our derivation of D_{ee}, using a mean field approach, cannot accurately describe the finer details when the endpoints search for each other over very small length scales, and hence, our theory must be modified in this regime.

We view the loop closure for small a ($<b$) as a two-step process (Figure 2), with the first being a reduction in $|\mathbf{R}_{ee}| \sim b$. The first stage is well-modeled by our modified SSS theory (see Figure 1A) using the effective diffusion coefficient in eq 12. The second stage involves a search for the two ends within a radius of b, so that contact can occur whenever $|\mathbf{R}_{ee}| = a < b$. The large-scale relaxations of the chain are not relevant in this regime. We therefore introduce a scale-dependent diffusion coefficient

$$D_{ee}(x) \approx \begin{cases} 8D_0/\sqrt{N\pi} & x > b \\ 2D_0 & x \leq b \end{cases} \tag{16}$$

Substitution of eq 16 into eq 6 with $P(r)$ given by eq 15 yields, for $a \leq b$,

$$\tau_c(a) \approx \frac{N^2b^2\pi}{24\sqrt{6}\,D_0} + \frac{N^{3/2}b^2(b-a)\sqrt{\pi}}{6\sqrt{6}\,D_0 a} \tag{17}$$

In Figure 1B, we compare the predictions of eq 17 for the

closure time to the simulated data for $a \leq b$. The fit is excellent, showing that the simple scale-dependent diffusion coefficient (eq 16), which captures the two-stage mechanism of cyclization when $a < b$, accurately describes the physics of loop closure for small a. By equating the two terms in eq 17, we predict that the $N^{3/2}$ scaling will begin to emerge when $N \lesssim 16b^2(a/b - 1)^2/a^2\pi$. This upper bound on N is consistent with the predictions of Chen et al.[30]

An alternate, but equivalent, description of the process of loop formation for small a can also be given. After the endpoints are within a sphere of radius b, chain fluctuations will drive them in and out of the sphere many times before contact is established. This allows us to describe the search process using an effective rate constant κ_{eff}, schematically shown in Figure 2. For small a, the loop closure (a search within radius b) becomes effectively rate-limited as opposed to diffusion-limited[35] contact formation. The search will be successful, in the SSS formalism, on a time scale of

$$\tau_{b\to a} \approx \frac{1}{2D_0 \mathcal{N}'} \int_a^b \frac{\mathrm{d}r}{P(r)} \left(\int_r^b \mathrm{d}r'\, P(r')\right)^2 \qquad (18)$$

with

$$\mathcal{N}' = \int_a^b \mathrm{d}r\, P(r)$$

Again, we have taken $D = 2D_0$ for $r < b$, because loop formation in this regime is dominated by the fast fluctuations of the monomers, which occurs on the time scale of b^2/D_0. For $a \approx b$, $\tau_{b\to a} \approx (a - b)^2/6D_0$, whereas $\tau_{b\to a} \approx b^3/6aD_0$ as $a \to 0$. The $\tau_{b\to a}$ can be used to define the effective rate constant $\kappa_{\text{eff}} \propto (b - a)/\tau_{b\to a}$. This can be substituted into eq 6, and gives the approximate loop closure time as $a \to 0$

$$\tau_c(a) - \tau_c(b) \approx \frac{1}{\kappa_{\text{eff}} \mathcal{N}' P(b)} \propto \frac{N^{3/2}b^3}{D_0 a} \qquad (19)$$

reproducing the same scaling for small a as that in eq 17.

The two-stage mechanism for the cyclization kinetics for $a/b < 1$ is reminiscent of the two-state kinetic mechanism used to analyze experimental data. The parameter κ_{eff} is analogous to the reaction-limited rate.[35] If the search rate within the capture region given by κ_{eff} is small, then we expect the exponent $\alpha_\tau < 2$. Indeed, the experiments of Buscaglia et al. suggest that α_τ changes from 2 (diffusion-limited) to 1.65 (reaction-limited). Our simulation results show the same behavior $\alpha_\tau = 2$ for $a/b \geq 1$, which corresponds to a diffusion-limited process, and $\alpha_\tau \approx 1.65$ for $a/b = 0.1$, in which the search within $a/b < 1$ becomes rate limiting.

3. Loop Closure for Polymers in Good and Poor Solvents

The kinetics of loop closure can change dramatically when interactions between monomers are taken into account. In good solvents, in which excluded volume interactions between the monomers dominate, it is suspected that only the scaling exponent in the dependence of τ_c on N changes compared to Rouse chains. However, relatively little is known about the kinetics of loop closure in poor solvents in which enthalpic effects, which drive collapse of the chain, dominate over chain entropy. Because analytic work is difficult when monomer−monomer interactions become relevant, we resort to simulations to provide insights into the loop closure dynamics.

3.1. Simulation of Cyclization Times. The Hamiltonian used in our simulations is $H = H_{\text{FENE}} + H_{\text{LJ}}$, where

$$H_{\text{FENE}} = -\frac{kb^2}{2} \sum_{i=1}^{N} \log\left[1 - \left(\frac{|\mathbf{r}_{i+1} - \mathbf{r}_i| - b}{R_0}\right)^2\right] \qquad (20)$$

models the chain connectivity, with $k = 22.2k_BT$ and $b = 0.38$ nm. The choice $R_0 = 2b/3$ (diverging at $|\mathbf{r}_{i+1} - \mathbf{r}_i| = b/3$ or $5b/3$) allowed for a larger time step than that when using[36] $R_0 = b/2$, and increased the efficiency of conformational sampling. The interactions between monomers are modeled using the Lennard-Jones potential

$$H_{\text{LJ}} = \epsilon_{\text{LJ}} \sum_{i=1}^{N-2} \sum_{j=i+2}^{N} \left[\left(\frac{b}{\mathbf{r}_{ij}}\right)^{12} - 2\left(\frac{b}{\mathbf{r}_{ij}}\right)^{6}\right] \qquad (21)$$

with $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$. The Lennard-Jones interaction between the covalently bonded beads $\mathbf{r}_i$ and $\mathbf{r}_{i+1}$ is neglected to avoid excessive repulsive forces. The second virial coefficient, defining the solvent quality, is given approximately by

$$\upsilon_2(\epsilon_{\text{LJ}}) = \int \mathrm{d}^3\mathbf{r}[1 - \exp(-\beta H_{\text{LJ}}(\mathbf{r}))] \qquad (22)$$

with $\beta = 1/k_BT$. In a good solvent $\upsilon_2 > 0$, while in a poor solvent $\upsilon_2 < 0$. A plot of υ_2 as a function of ϵ_{LJ} given in Figure 3A shows that $\upsilon_2 > 0$ when $\beta\epsilon_{\text{LJ}} < 0.3$ and $\upsilon_2 < 0$ if $\beta\epsilon_{\text{LJ}} > 0.3$. In what follows, we will refer to $\beta\epsilon_{\text{LJ}} = 0.4$ as weakly hydrophobic and $\beta\epsilon_{\text{LJ}} = 1.0$ as strongly hydrophobic. The classification of the solvent quality based on eq 22 is approximate. The precise determination of the Θ point ($\upsilon_2 \approx 0$) requires the computation of υ_2 for the entire chain. For our purposes, this approximate demarcation between good, Θ, and poor solvents based on eq 22 suffices.

To fully understand the effect of solvent quality on the cyclization time, we performed Brownian dynamics simulations for $\beta\epsilon_{\text{LJ}} = i/10$, with $1 \leq i \leq 10$. In our simulations, N was varied from 7 to 300 for each value of ϵ_{LJ}, with a fixed capture radius of $a = 2b = 0.76$ nm. The loop closure time was identified with the mean first passage time. The dynamics for each trajectory was followed until the two ends were within the capture radius a. Averaging the first passage times over 400 independent trajectories yielded the mean first passage time. The chains were initially equilibrated using parallel tempering (replica exchange) Monte Carlo[37] to ensure proper equilibration, with each replica pertaining to one value of ϵ_{LJ}. In Figure 3B, we show the scaling of the radius of gyration $\langle \mathbf{R}_g^2 \rangle$ as a function of N. We find $\langle \mathbf{R}_{ee}^2 \rangle \sim N^{6/5}$ for the good solvent and $\langle \mathbf{R}_{ee}^2 \rangle \sim N$ for the Θ solvent ($\beta\epsilon_{\text{LJ}} = 0.3$). In poor solvents ($\beta\epsilon_{\text{LJ}} > 0.3$), the large N scaling of $\langle \mathbf{R}_{ee}^2 \rangle \sim N^{2/3}$ is not observed for the values of N used in our simulations. Similar deviations from the expected scaling of $\langle \mathbf{R}_{ee}^2 \rangle$ with N have been observed by Rissanou et al.[38] for short chains in a poor solvent. Simulations using much longer chains ($N \gtrsim 5000$) may be required to observe the expected scaling exponent of 2/3.

Brownian dynamics simulations with $D_0 = 0.77$ nm^2/ns ($= k_BT/6\pi\eta b$, with $\eta = 1.5$ cP) were performed to determine τ_c. The loop closure time for the chains in varying solvent conditions is shown in Figure 4A and B. The solvent quality drastically changes the loop closure time. The values of τ_c for the good solvent ($\beta\epsilon_{\text{LJ}} = 0.1$) are nearly 3 orders of magnitude larger than those in the case of the strong hydrophobe ($\beta\epsilon_{\text{LJ}} = 1.0$) for $N = 80$ (Figure 4A). For N in the range of 20−30, which are typically used in experiments on tertiary contact formation in polypeptide chains, the value of τ_c is about 20 ns in good solvents, whereas in poor solvents, τ_c is only about 0.3 ns. The results are vividly illustrated in Figure 4B, which shows

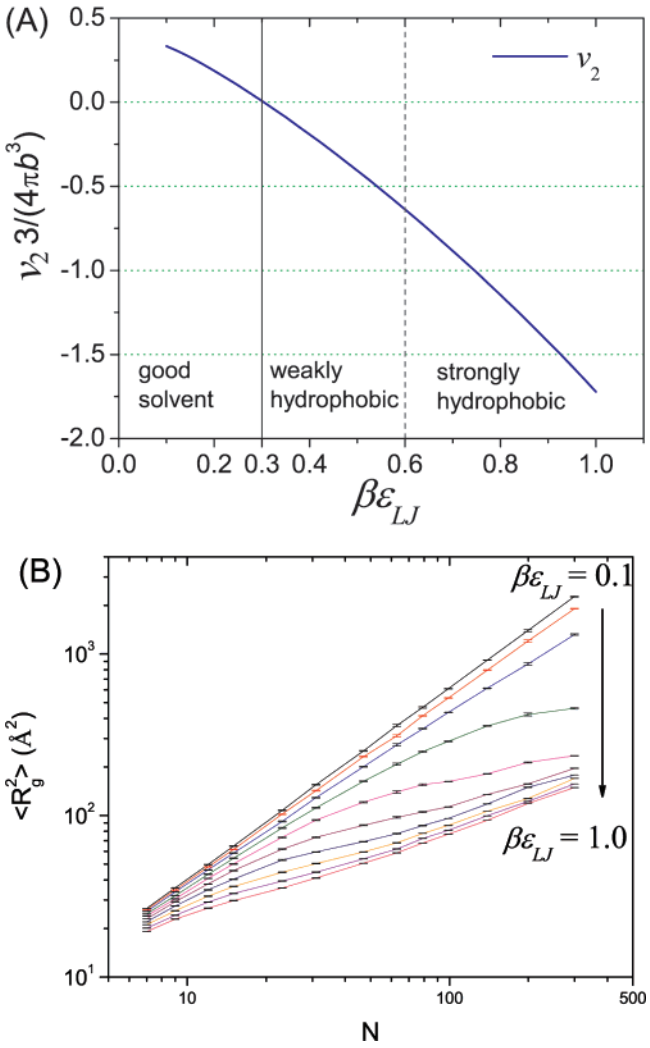

Figure 3. (A) Second virial coefficient as a function of ϵ_{LJ} from eq 22. The classification of solvent quality based on the values of v_2 are shown. (B) The variation of $\langle \mathbf{R}_g^2 \rangle$ with N for different values of ϵ_{LJ}. The value of $\beta\epsilon_{LJ}$ increases from 0.1 to 1.0 (in the direction of the arrow).

τ_c as a function of ϵ_{LJ} for various N values. The differences in τ_c are less pronounced as N decreases (Figure 4B). The absolute value of τ_c for $N \approx 20$ is an order of magnitude less than that obtained for τ_c in polypeptides.[35] There could be two interrelated reasons for this discrepancy. The value of D_0, an effective diffusion constant in the SSS theory, extracted from experimental data and simulated $P(\mathbf{R}_{ee})$ is about an order of magnitude less than the D_0 in our paper. Second, Buscaglia et al.[35] used the WLC model with excluded volume interactions, whereas our model does not take into account the effect of bending rigidity. Indeed, we had shown in an earlier study[33] that chain stiffness increases τ_c. Despite these reservations, our values of τ_c can be made to agree better with experiments using $\eta \approx 5$ cP[9] and a slightly larger value of b. Because it is not our purpose to quantitatively analyze cyclization kinetics in polypeptide chains, we did not perform such a comparison.

We also find that the solvent quality significantly changes the scaling of $\tau_c \sim N^{\alpha_\tau}$, as shown in Figure 4C. For the range of N considered in our simulations, τ_c does not appear to vary as a simple power law in N (much like $\langle \mathbf{R}_g^2 \rangle$; see Figure 3B) for $\beta\epsilon_{LJ} > 0.3$. The values of τ_c in poor solvents shows increasing curvature as N increases. However, if we insist that a simple power law describes the data then for the smaller range of N from 7 to 32 (consistent with the methods of other au-

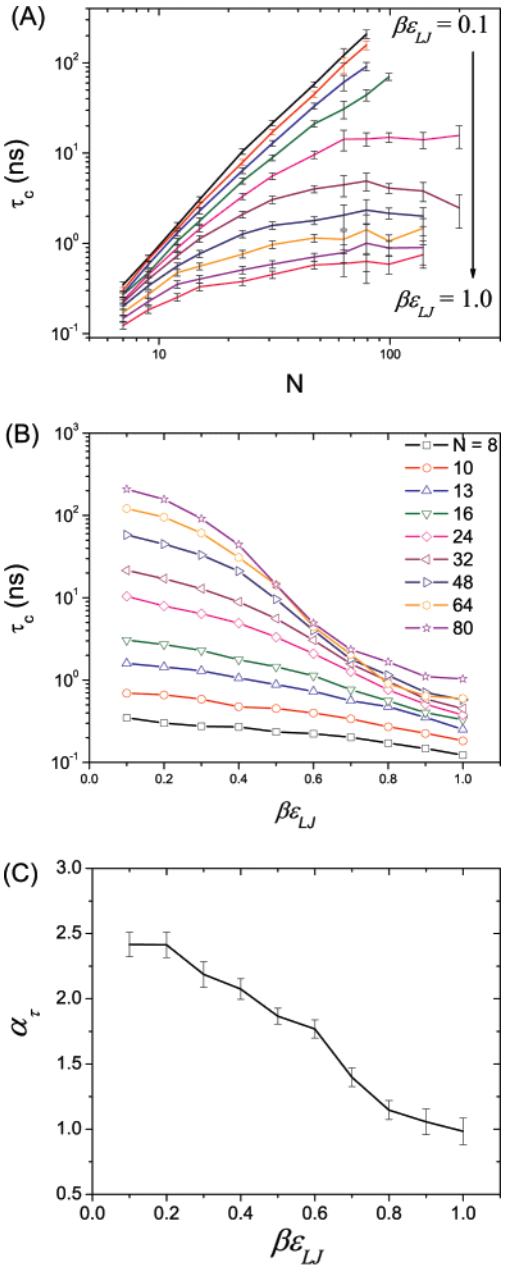

Figure 4. (A) Loop closure time as a function of N for varying solvent quality. The values of $\beta\epsilon_{LJ}$ increase from 0.1 to 1.0 from top to bottom, as in Figure 3A. (B) τ_c as a function of ϵ_{LJ}, which is a measure of the solvent quality. The values of N are shown with various symbols. (C) Variation of the scaling exponent of $\tau_c \sim N^{\alpha_\tau}$ as a function of ϵ_{LJ}.

thors[16,22,35]), we can fit the initial slopes of the curves to determine an effective exponent α_τ (Figure 4C), that is, $\tau_c \approx \tau_0 N^{\alpha_\tau}$. In the absence of sound analytical theory, the extracted values of α_τ should be viewed as an effective exponent. We anticipate that, much like the scaling laws for $\langle \mathbf{R}_g^2 \rangle$, the final large N scaling exponent for τ_c will only emerge for[38] $N \gtrsim 5000$, which is too large for accurate simulations. However, with the assumption of a simple power law behavior for small N, we find that the scaling exponent precipitously drops from $\alpha_\tau \approx 2.4$ in the good solvent to 1.0 in the poor solvent. Our estimate of α_τ in good solvents is in agreement with the prediction of Debnath and Cherayil[7] ($\alpha_\tau \approx 2.3-2.4$) or Thirumalai[39] ($\alpha_\tau \approx 2.4$) and is fairly close to the value obtained in previous simulations[31] ($\alpha_\tau \approx 2.2$). The difference in the scaling exponent between the present and previous study[31] may be related to the choice of the Hamiltonian in the simulations. Podtelezhnikov

and Vologodskii[31] used a harmonic repulsion between monomers to represent the impenetrability of the chain and took $a/b < 1$ in their simulations. Scaling arguments predict $\alpha_\tau = 2.2$ for a sufficiently long chain in good solvent (see appendix A), suggesting that our value of $\alpha_\tau \approx 2.4$ may be due to finite size effects.

In contrast to the good solvent case, our estimate of α_τ in poor solvents is significantly lower than the predictions of Debnath and Cherayil,[7] who suggested $\alpha_\tau \approx 1.6-1.7$ based on a modification of the WF formalism.[3] However, fluorescence experiments on multiple repeats of the possibly weakly hydrophobic glycine and serine residues in D_2O have found $\tau_c \sim N^{1.36}$ for short chains[22] and $\tau_c \sim N^{1.05}$ for longer chains,[16] in qualitative agreement with our simulation results. Bending stiffness[26,33] and hydrodynamic interactions may make direct comparison between these experiments and our results difficult. The qualitative agreement between simulations and experiments on polypeptide chains suggests that interactions between monomers are more important than hydrodynamic interactions, which are screened.

3.2. Mechanisms of Loop Closure in Poor Solvents. The dramatically smaller loop closure times in poor solvents than those in good solvents (especially for $N > 20$; see Figure 4B) requires an explanation. In poor solvents, the chain adopts a globular conformation with the monomer density of $\rho b^3 \sim O(1)$, where $\rho \approx N/R_g^3$. We expect the motions of the monomers to be suppressed in the dense, compact globule. For large N, when entanglement effects may dominate, it could be argued that in order for the initially spatially separated chain ends ($|\mathbf{R}_{ee}|/a > 1$) to meet, contacts between the monomer ends with their neighbors must be broken. Such unfavorable events might require overcoming enthalpic barriers ($\approx \bar{Q} \times \epsilon_{LJ}$, where $\bar{Q}$ is the average number of contacts for a bead in the interior of the globule), which would increase τ_c. Alternatively, if the ends search for each other using a diffusive, reptation-like mechanism without having to dramatically alter the global shape of the collapsed globule, τ_c might decrease as ϵ_{LJ} increases (i.e., as the globule becomes more compact). It is then of interest to ask whether looping events are preceded by global conformational changes, with a large-scale expansion of the polymer that allows the endpoints to search the volume more freely, or if the endpoints search for each other in a highly compact, but more restrictive, ensemble of conformations.

In order to understand the mechanism of looping in poor solvents, we analyze in detail the end-to-end distance $|\mathbf{R}_{ee}(t)|$ and the radius of gyration $|\mathbf{R}_g(t)|$ for two trajectories (with $\beta\epsilon_{LJ} = 1$ and $N = 100$). One of the trajectories has a fast looping time ($\tau_c^F \approx 0.003$ ns), while the looping time in the other is considerably slower ($\tau_c^S \approx 4.75$ ns). Additionally, we compute the time-dependent variations of the coordination number, $Q(t)$, for each endpoint. We define two monomers i and j to be in "contact" if $|\mathbf{r}_i - \mathbf{r}_j| \geq 1.23b$ (beyond which the interaction energy $E_{LJ} \geq -\epsilon_{LJ}/2$) and define $Q_1(t)$ and $Q_N(t)$ to be the total number of monomers in contact with monomers 1 and N, respectively. We do not include nearest neighbors on the backbone when computing the coordination number, and the geometrical constraints give $0 \leq Q(t) \leq 11$ for either endpoint. With this definition, an endpoint on the surface of the globule will have $Q = 5$. These quantities are shown in Figures 5 and 6.

The trajectory with τ_c^F (Figure 5) shows little variation in either $|\mathbf{R}_g|$ or $|\mathbf{R}_{ee}|$. We find $|\mathbf{R}_{ee}| \approx |\mathbf{R}_g|$, suggesting that the endpoints remain confined within the dense globular structure throughout the looping process. This is also reflected in the coordination numbers for both of the endpoints, with both $Q_1(t)$ and $Q_N(t)$ in the range of $5 \leq Q(t) \leq 10$ throughout the simulation. The endpoints in this trajectory, with the small loop closure time τ_c^F, always have a significant number of contacts and traverse the interior of the globule when searching for each other. Similarly, we also found that the trajectory with a long first passage time τ_c^S (Figure 6) shows little variation in $\mathbf{R}_g$ throughout the run. The end-to-end distance, however, shows large fluctuations over time, and $\langle \mathbf{R}_{ee}^2 \rangle^{1/2} \gtrsim 2\langle \mathbf{R}_g^2 \rangle^{1/2}$ until closure. This suggests that, while the chain is in an overall globular conformation (small, constant $\mathbf{R}_g^2$), the endpoints are mainly found on the exterior of the globule. This conclusion is again supported by the coordination number, with $Q(t) \leq 5$ for significant portions of the simulation. While the endpoints are

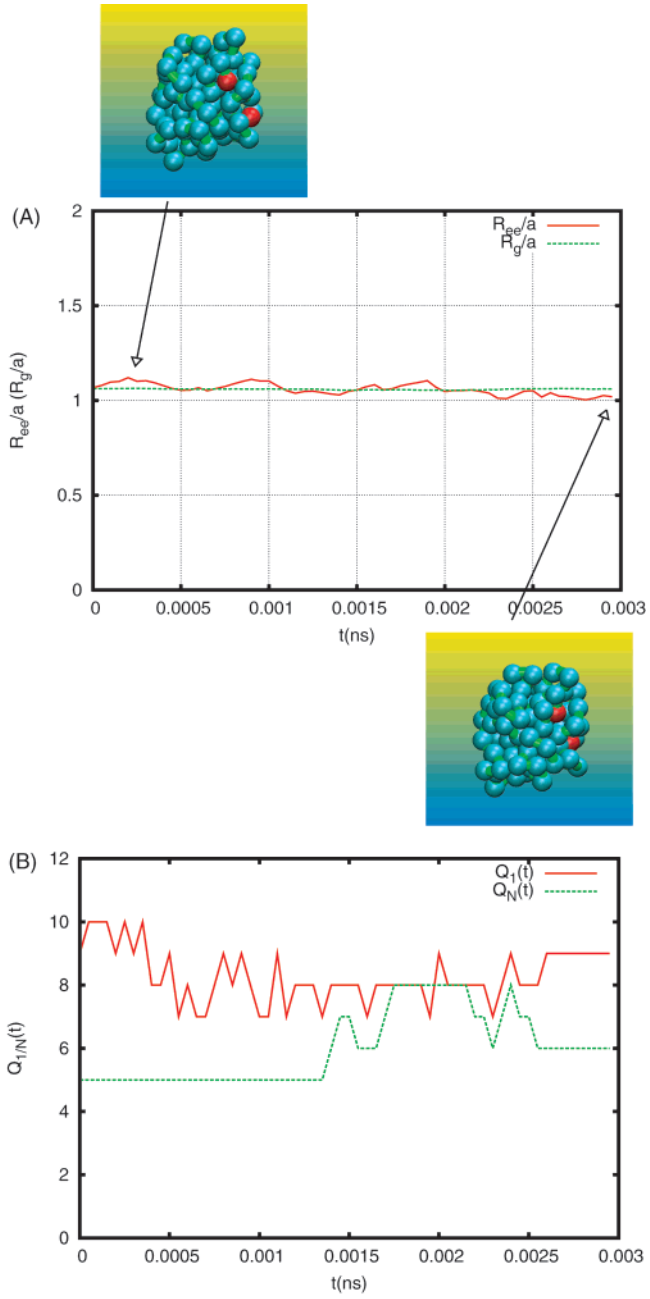

Figure 5. Mechanism of loop closure for a trajectory with a short (~0.003 ns) first passage time. The values of N and $\beta\epsilon_{LJ}$ are 100 and 1.0, respectively. (A) Plots of $|\mathbf{R}_{ee}|$ and $|\mathbf{R}_g|$ (scaled by the capture radius a) as a function of time. The structures of the globules near the initial stage and upon contact formation between the ends are shown. The end-to-end distance is in red. (B) The time-dependent changes in the coordination numbers for the first ($Q_1(t)$) and last ($Q_N(t)$) monomers during the contact formation.

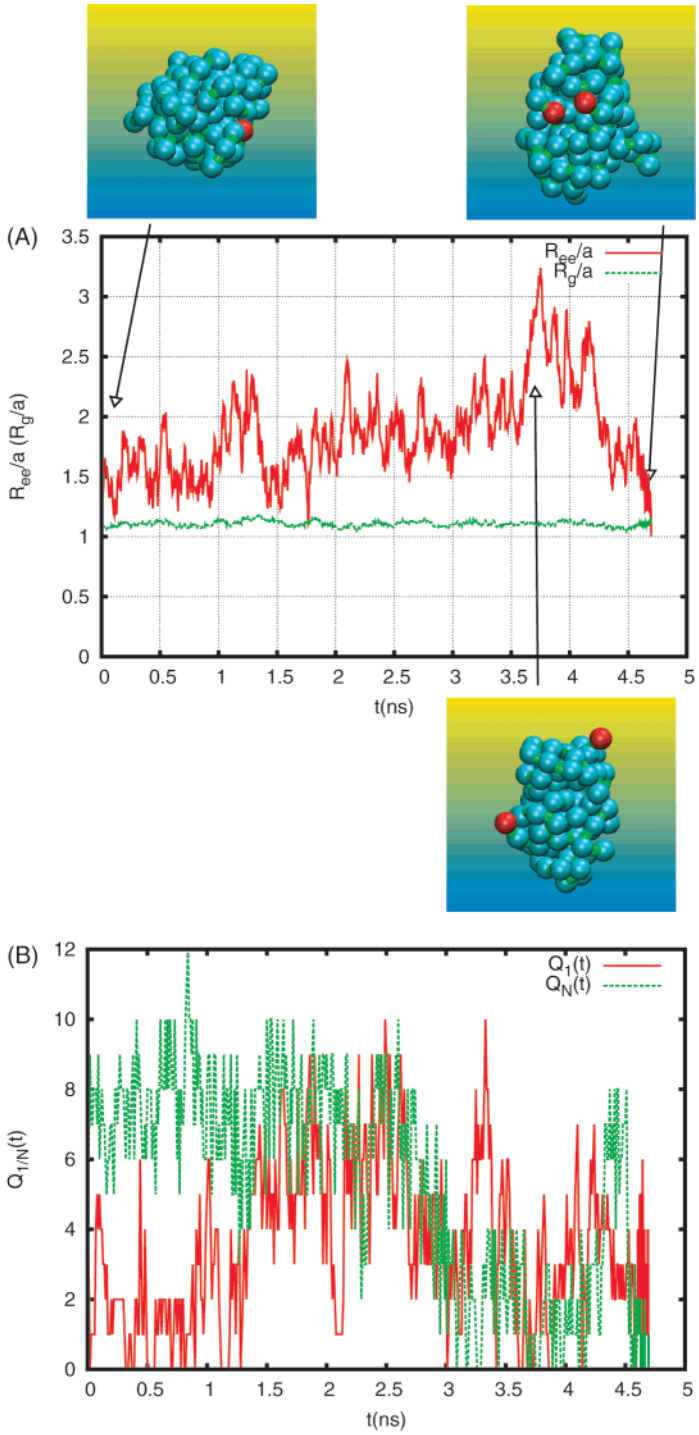

Figure 6. Same as Figure 5, except the data are for a trajectory with a first passage time for contact formation that is about 4.7 ns. (A) Although the values of $|\mathbf{R}_g|$ are approximately constant, $|\mathbf{R}_{ee}|$ fluctuates greatly. (B) Substantial variations in $Q_1(t)$ and $Q_N(t)$ are observed during the looping dynamics, in which both ends spend a great deal of time on the surface of the globule.

less restricted by nearby contacts and able to fluctuate more, they spend a much longer time searching for each other. Thus, it appears that the process of loop formation in poor solvents, where enthalpic effects might be expected to dominate for $N = 100$, occurs by a diffusive, reptation-like process. Entanglement effects are not significant in our simulations.

We note that trajectories in which the first passage time for looping is rapid (with $\tau_c^i < \tau_c$ for trajectory i) have at least one endpoint with a high coordination number ($Q > 5$) throughout the simulation. In contrast, for most slow-looping runs (with $\tau_c^i > \tau_c$), we observe long stretches of time where both endpoints have a low coordination number ($Q < 5$). These results suggest that motions within the globule are far less restricted than one might have thought, and loop formation will occur faster when the endpoints are within the globule than it would if the endpoints were on the surface. The longer values of τ_c are found if the initial separation of the endpoints is large, which is more likely if they are on the surface than if they are buried in the interior. The absence of any change in $|\mathbf{R}_g(t)|$ in both the trajectories, which represent the extreme limits in the first passage time for looping, clearly shows that contact formation in the globular phase is not an activated process. Thus, we

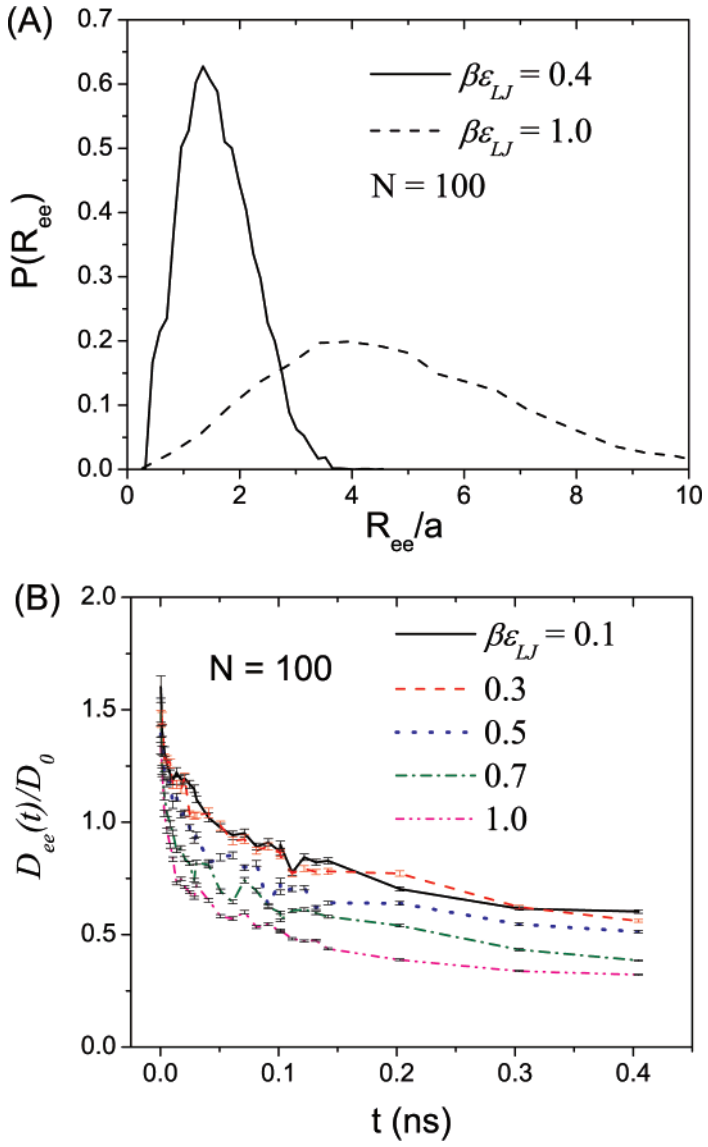

Figure 7. (A) Distribution of end-to-end distances for a weakly ($\beta\epsilon_{LJ} = 0.4$) and strongly ($\beta\epsilon_{LJ} = 1.0$) hydrophobic chain. (B) Diffusion constant $D_{ee}(t)$ in units of D_0 for varying solvent quality. The diffusion constant is defined using $D_{ee}(t) = \langle \delta \mathbf{R}_{ee}^2(t) \rangle / 6t$. The values of ϵ_{LJ} are shown in the inset.

surmise that looping in poor solvents occurs by a diffusive, reptation-like mechanism, provided entanglement effects are negligible.

3.3. Separating the Equilibrium Distribution $P(\mathbf{R}_{ee})$ and Diffusive Processes in Looping Dynamics. The results in the previous section suggest a very general mechanism of loop closure for interacting chains. The process of contact formation for a given trajectory depends on the initial separation $\mathbf{R}_{ee}$, and the dynamics of the approach of the ends. Thus, τ_c should be determined by the distribution of $P(\mathbf{R}_{ee})$ (an equilibrium property) and an effective diffusion coefficient $D(t)$ (a dynamic property). We have shown for the Rouse model that such a deconvolution into equilibrium and dynamical parts, which is in the spirit of the SSS approximation, is accurate in obtaining τ_c for a wide range of N and a/b. It turns out that a similar approach is applicable to interacting chains as well.

The decomposition of looping mechanisms into a convolution of equilibrium and dynamical parts explains the large differences in τ_c as the solvent quality changes. We find, in fact, that the equilibrium behavior of the endpoints dominates the process of loop formation, with the kinetic processes being only weakly dependent on the solvent quality for short chains. In Figure 7A, we plot the end-to-end distribution function for weakly ($\beta\epsilon_{LJ} = 0.4$) and strongly ($\beta\epsilon_{LJ} = 1$) hydrophobic polymer chains. The strongly hydrophobic chain is highly compact, with a sharply peaked distribution. The average end-to-end distance is significantly lower than is the weakly hydrophobic case. While the distribution function is clearly strongly dependent on the interactions, the diffusion coefficient $D(t)$ is only weakly dependent on the solvent quality (Figure 7B). The values of $D(t) = \langle \delta \mathbf{R}_{ee}^2 \rangle / 6t$ are only reduced by a factor of about 2 between the $\beta\epsilon_{LJ} = 0.1$ (good solvent, with a globally swollen configuration) and the $\beta\epsilon_{LJ} = 1.0$ (poor solvent, with a globally globular configuration) on intermediate time scales. We note, in fact, that the good solvent and Θ solvent cases have virtually identical diffusion coefficients throughout the simulations (Figure 7B). This suggests that the increase in τ_c (Figure 4) between the Rouse chains and the good solvent chains is primarily due to the broadening of the distribution $P(\mathbf{R}_{ee})$, that is, the significant increase in the average end-to-end distance in the good solvent case, $\langle \mathbf{R}_{ee}^2 \rangle \sim N^{2\nu}$, with $\nu = 3/5$.

Because of the weak dependence of the diffusion coefficient on the solvent quality, the loop closure time is dominated primarily by the end-to-end distribution function. In other words, the equilibrium distribution function $P(\mathbf{R}_{ee})$, to a large extent, determines τ_c. To further illustrate these arguments, we find that if we take $D \approx 2D_0$ in eq 6 and numerically integrate the distribution function found in the simulations for $N = 100$, $\tau_c(\beta\epsilon_{LJ} = 1.0)$ and $\tau_c(\beta\epsilon_{LJ} = 0.4)$ differ by 2 orders of magnitude, almost completely accounting for the large differences seen in Figure 4B between the two cases. Moreover, if the numerically computed values of $D(t)$ for long t ($t > 0.5$ ns in Figure 7, for example) are used for D_{ee} in eq 6, we obtain values of τ_c that are in reasonably good agreement with simulations. The use of D_{ee} ensures that the dynamics of the entire chain is explicitly taken into account. These observations rationalize the use of $P(\mathbf{R}_{ee})$ with a suitable choice of D_{ee} in obtaining accurate results for flexible as well as stiff chains.[33,40] Because $P(\mathbf{R}_{ee})$ can, in principle, be inferred from FRET experiments,[41,42] the theory outlined here can be used to quantitatively predict loop formation times. In addition, FRET experiments can also be used to assess the utility of polymer models in describing fluctuations in single-stranded nucleic acids and polypeptide chains.

3.4. Kinetics of Interior Loop Formation. We computed the kinetics of contact between beads that are in the chain interior as a function of solvent quality (Figure 8A) using $N = 32$. The mean time for making a contact is computed using the same procedure as that used for cyclization kinetics. For simplicity, we only consider interior points that are centered around the midpoint of the chain. The ratio r_l, which measures the change in the time for interior loop formation relative to cyclization kinetics, depends on $\beta\epsilon_{LJ}$ and l/N, where l is the separation between the beads (Figure 8A). The nonmonotonic dependence of r_l on l in good solvents further shows that as l/N decreases to about 0.6, $r_l \approx 1$. The maximum in r_l at $l/N \approx 0.9$ decreases as $\beta\epsilon_{LJ}$ increases. In the poorest solvents considered ($\beta\epsilon_{LJ} = 0.8$), we observe that r_l only decreases monotonically with decreasing l/N. Interestingly, in poor solvents, r_l can be much less than unity, which implies that it is easier to establish contacts between beads in the chain interior than between the ends. This prediction can be verified in polypeptide chains in the presence of inert crowding agents that should decrease the solvent quality. Just as in cyclization kinetics, interior loop formation also depends on the interplay between internal chain diffusion that gets slower as the solvent quality decreases and equilibrium distribution (which gets narrower) of the distance between the contacting beads.

We also performed simulations for $N = 80$ by first computing the time for cyclization τ_c^{80}. In another set of simulations, two flexible linkers, each containing 20 beads, were attached to the ends of the $N = 80$ chain. For the resulting longer chain, we calculated τ_l for $l = 80$ as a function of $\beta\epsilon_{LJ}$. Such a calculation is relevant in the context of single-molecule experiments in which the properties of a biomolecule (RNA) is inferred by attaching linkers with varying polymer characteristics. It is important to choose the linker characteristics that minimally affect the dynamic properties of the molecule of interest. The ratio $\tau_{l=80}/\tau_c^{80}$ depends on $\beta\epsilon_{LJ}$ and changes from 2.6 (good solvents) to 2.0 under the Θ condition and becomes unity in poor solvents (Figure 8B). Analysis of the dependence of the diffusion coefficients of interior-to-interior vector D_{ij} ($i = 20$ and $j = 100$) and end-to-end vector (of the original chain without linkers) D_{ee} on solvent conditions indicates that on the time scales relevant to the loop closure time (analogous to τ_{ee} for the Rouse chain), D_{ij} reduces to about half of D_{ee} in good and Θ solvents, whereas the two are very similar in poor solvents. The changes in the diffusion coefficient together with the equilibrium distance distribution explains the behavior in Figure 8B.

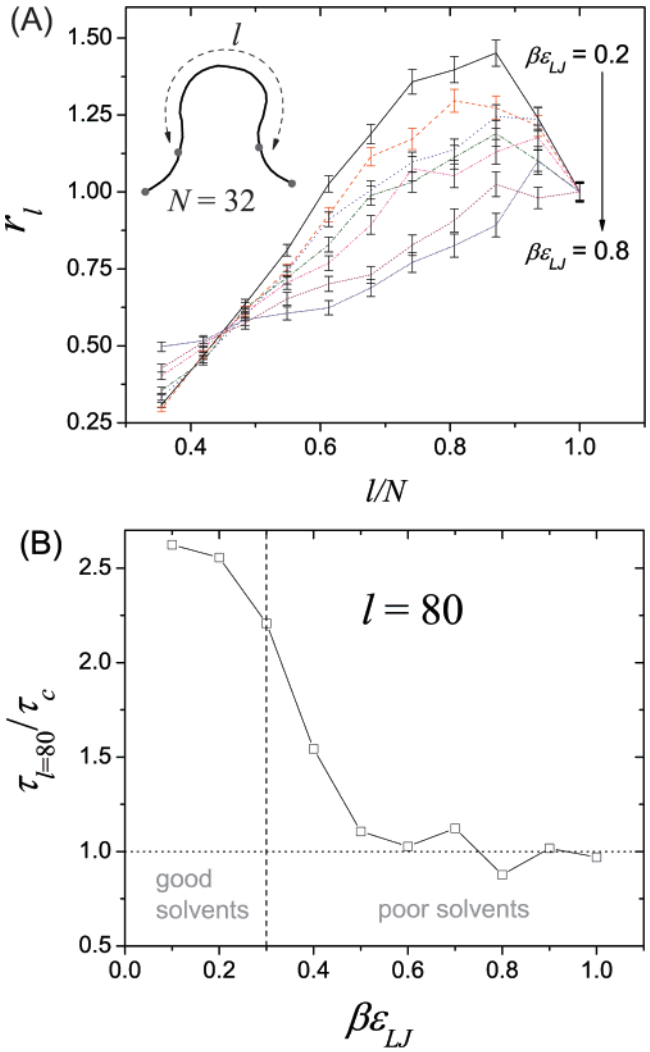

Figure 8. (A) The ratio $r_l = \tau_l/\tau_c$ as a function of interior length l. Here, τ_l is the contact formation time for beads that are separated by l monomers; r_l is nonmonotonic for weakly hydrophobic chains but increases monotonically with l in the poorest solvents. The observed maxima occur near $l/N = 0.9$. (B) For loop length $l = 80$, the ratio $\tau_{l=80}/\tau_c$ as a function of $\beta\epsilon_{LJ}$ for a chain with two linkers (each of 20 beads) that are attached to beads 20 and 100. In good solvents, the interior loop closure kinetics is about 2.5 times slower than the end-to-end one with the same loop length. In poor solvents, however, there is virtually no difference between the two.

4. Conclusions

A theoretical description of contact formation between the chain endpoints is difficult because of the many-body nature of the dynamics of a polymer. Even for the simple case of cyclization kinetics in Rouse chains, accurate results for τ_c are difficult to obtain for all values of N, a, and b. The present work confirms that, for large N and $a/b > 1$, the looping time must scale as N^2, a result that was obtained some time ago using the WF formalism.[3,5] Here, we have derived $\tau_c \sim N^2$ (for $N \gg 1$ and $a \geq b$) by including the full internal chain dynamics within the simple and elegant SSS theory.[4] We have shown that, for $N < 100$ and especially in the (unphysical) limit $a/b < 1$, the loop

closure time is $\tau_c \sim \tau_0 N^{\alpha_\tau}$, with $1.5 < \alpha_\tau \leq 2$. In this limit, our simulations show that loop closure occurs in two stages with vastly differing time scales. By incorporating these processes into a scale-dependent diffusion coefficient, we obtain an expression for τ_c that accurately fits the simulation data. The resulting expression for τ_c for $a < b$ (eq 17) contains both the $N^{3/2}$ and N^2 limits, as was suggested by Pastor et al.[28]

The values of τ_c for all N change dramatically when interactions between monomers are taken into account. In good solvents, $\tau_c \sim \tau_0 N^{\alpha_\tau}$ ($\alpha_\tau \approx 2.4$) in the range of N used in the simulations. Our exponent α_τ is in reasonable agreement with earlier theoretical estimates.[7,39] Polypeptide chains in high denaturant concentrations may be modeled as flexible chains in good solvents. From this perspective, the simple scaling law can be used to fit the experimental data on loop formation in the presence of denaturants using physical values of τ_0. Only when N is relatively small ($N \approx 4$) will chain stiffness play a role in controlling loop closure times. Indeed, experiments show that τ_c increases for short N (see Figure 3 in ref 15) and deviates from the power law behavior given in eq 7 for all N, which is surely due to the importance of bending rigidity.

The simulation results for τ_c in poor solvents show rich behavior that reflects the extent to which the quality of the solvent is poor. The poorness of the solvent can be expressed in terms of

$$\lambda = \frac{\epsilon_{LJ} - \epsilon_{LJ}(\Theta)}{\epsilon_{LJ}(\Theta)} \tag{23}$$

where the Θ solvent interaction strength $\beta\epsilon_{LJ}(\Theta) \approx 0.3$ is determined from $\upsilon_2 \approx 0$ (Figure 3). Loop closure times decrease dramatically as λ increases. For example, τ_c decreases by a factor of about 100 for $N = 80$ as λ increases from 0 to 2.3. In this range of N, a power law fit of τ_c with N ($\tau_c \sim N^{\alpha_\tau}$) shows that the exponent α_τ depends of λ. Analysis of the trajectories that monitor loop closure shows that contact between each end of the chains is established by mutual, reptation-like motion within the dense, compact globular phase.

The large variations of τ_c as λ changes suggest that there should be significant dependence of the loop formation rates on the sequence in polypeptide chains. In particular, our results suggest that as the number of hydrophobic residues increase, τ_c should decrease. Similarly, as the number of charged or polar residues increase, the effective persistence length (l_p) and interactions can be altered, which in turn could increase τ_c. Larger variations in τ_c, due to its dependence on l_p and N, can be achieved most easily in single-stranded RNA and DNA. These arguments neglect sequence effects, which are also likely to be important. The results in Figure 4B may also be reminiscent of "hydrophobic collapse" in proteins, especially as λ becomes large. For large λ and long N, it is likely that τ_c correlates well with time scales for collapse. This scenario is already reflected in $P(\mathbf{R}_{ee})$ (see Figure 7A). It may be possible to discern the predictions in Figure 4B by varying the solvent quality for polypeptides. A combination of denaturants (makes the solvent quality good) and PEG (makes it poor) can be used to measured τ_c in polypeptide chains. We expect that the measured τ_c should be qualitatively similar to the findings in Figure 4B.

The physics of loop closure for small and intermediate chain lengths ($N \leq 300$) is rather complicated due to contributions from various time and length scales (global relaxation and internal motions of the chains). The contributions from these sources are often comparable, making the process of looping dynamics difficult to describe theoretically. A clear picture of the physics is obtained only when one considers all possible ranges of the parameters entering the loop closure time equation. To this end, we have explored wide ranges of conceivable parameters, namely, the chain length N, capture radius a, and conditions of the solvents expressed in terms of ϵ_{LJ}. By combining analytic theory and simulations, we have shown that, for a given N, the looping dynamics in all solvent conditions is primarily determined by the initial separation of the endpoints. The many-body nature of the diffusive process is embodied in $D(t)$, which does not vary significantly as λ changes for a fixed N. Finally, the dramatic change in τ_c as λ increases suggests that it may be also necessary to include hydrodynamic interactions, which may decrease τ_c further, to more accurately obtain the loop closure times.

Acknowledgment. This work was supported, in part, by a grant from the National Science Foundation through Grant Number NSF CHE 05-14056.

5. Appendix A

Friedman and O'Shaughnessy[43] (FO) generalized the concept of the exploration of space suggested by de Gennes[44] to the cyclization reaction of polymer chains. The arguments given by de Gennes and FO succinctly reveal the conditions under which local equilibrium is appropriate in terms of properties of the polymer chains.

First, de Gennes introduced the notion of compact and noncompact exploration of space associated with a bimolecular reaction involving polymers. Tertiary contact formation is a particular example of such a process. Consider the relative position between two reactants on a lattice with the lattice spacing a. The two reactants explore the available conformational space until their relative distance becomes less than the reaction radius. One can define two quantities relevant to the volume spanned prior to the reaction. One comes from the actual number of jumps on the lattice defined as $j(t)$, which is directly proportional to t. If the jump is performed in a d-dimensional lattice, the actual volume explored would be $a^d j(t)$. The other quantity comes from the root-mean-square distance. If $x(t) \sim t^u$ is the root-mean-square distance for one-dimension, $x^d(t)$ is the net volume explored. The comparison between these two volumes defines the compactness in the exploration of the space. (i) The case $x^d(t) > a^d j(t)$ corresponds to noncompact exploration of the space ($ud > 1$). (ii) The regime $x^d(t) < a^d j(t)$ represents compact exploration of the space ($ud < 1$). Depending on the dimensionality, the exploration of space by the reactive pair in the bimolecular reaction is categorized either into noncompact ($d = 3$) or into compact ($d = 1$) exploration. In the case of noncompact exploration, the bimolecular reaction takes place infrequently, so that the local equilibrium in solution is easily reached. The reaction rate is simply proportional to the probability that the reactive pair is within the reaction radius, so that $k \sim p_{eq}(r < r_0)$, which eventually leads to $k = 4\pi\sigma D$, the well-known steady-state diffusion-controlled rate coefficient. It can be shown that $k \sim t^{ud-1}$ in the case of compact exploration.

In the context of polymer cyclization, the compactness of the exploration of space can be assessed using the exponent $\theta = (d + g)/z$, where g is the correlation hole exponent and z is the dynamic exponent, such that $r \sim t^{1/z}$. Since[45,46]

$$\lim_{r \to 0} p_{eq}(r) \sim \frac{1}{R^d}\left(\frac{r}{R}\right)^g$$

where $R = \mathbf{R}_{ee}$. The cyclization rate can be approximated by $k \sim (\mathrm{d}/\mathrm{d}t) \int d^d r p_{eq}(r)$, and it follows that $k \sim (\mathrm{d}/\mathrm{d}t)(r/R)^{d+g}$. The relations $r \sim t^{1/z}$ and $R \sim \tau^{1/z}$ lead to $k \sim (1/\tau)(t/\tau)^{((d+g)/z)-1}$, where τ is the characteristic relaxation time.

(1) If $\theta > 1$, then the cyclization rate is given by $k \sim p_{eq}(r = r_0) \sim (1/R^d) f(r/R)$. which, with $R \sim N^\nu$, leads to the scaling relation

$$\tau_c \sim N^{\nu(d+g)} \tag{24}$$

(2) If $\theta < 1$, the compact exploration of conformations occurs between the chain ends. As a result, the internal modes are not in local equilibrium. In this case, $\tau_c \sim \tau_R \sim R^z$, where $z = 2 + (1/\nu)$ is the dynamic exponent for free-draining case and $z = d$ when hydrodynamic interactions are included.[34,45] Therefore, the scaling law for the cyclization rate is given by

$$\tau_c \sim N^{z\nu} \tag{25}$$

The inference about the validity of local equilibrium, based on θ, is extremely useful in obtaining the scaling laws for polymer cyclization, eqs 24 and 25. Extensive Brownian dynamics simulation by Rey et al.[47] have established the validity of these scaling laws. The expected scaling laws for three different polymer models are discussed below.

Free-Draining Gaussian Chain ($d = 3$, $g = 0$, $z = 4$, $\nu = (1/2)$); $\theta = 3/4 < 1$. Because $\theta < 1$, the local equilibrium approximation is not valid for a "long" free-draining Gaussian chain or, equivalently, the Rouse model. Accordingly, we expect $\tau_c \sim N^2$ for the Rouse chain for $N \gg 1$. However, if N is small and the local equilibrium is established among the internal Rouse modes so that $\tau_c \gg \tau_R$, the scaling relation change from $\tau_c \sim N^2$ to $\sim N^{\alpha_\tau}$, with $\alpha_\tau < 2$. The simulations shown here and elsewhere[28] and the theory by Sokolov[6] explicitly demonstrate that α_τ can be less than 2 for small N. In this sense, the looping time of the free-draining Gaussian chain of finite size is bound by[25,33] $\tau_{SSS} < \tau_c < \tau_{WF}$.

Free-Draining Gaussian Chain with Excluded Volume ($d = 3$, $g = (\gamma - 1)/\nu = 5/18$, $z = 11/3$, $\nu = 3/5$); $\theta = 59/66 < 1$. From eq 25, it follows that $\tau_c \sim N^{2.2}$. This polymer model has been extensively studied using Brownian dynamics simulation, and the value of the scaling exponent 2.2 has been confirmed by Vologodskii.[31] The value of the exponent (2.2) is also consistent with previous theoretical predictions.[7,39]

Gaussian Chain with Excluded Volume and Hydrodynamic Interactions ($d = 3$, $g = 5/18$, $z = 3$, $\nu = 3/5$); $\theta = 59/54 > 1$. Since $\theta > 1$, the local equilibrium approximation is expected to hold. This polymer model corresponds to the flexible polymer in a good solvent. The incorporation of hydrodynamic interactions may assist the fast relaxation of the rapid internal modes and changes the nature of the cyclization dynamics from a compact to a noncompact one. The correct scaling law is predicted to be $\tau_c \sim N^{2.0}$. Since the local equilibrium approximation is correct, the first passage time approach[4] should give a correct estimate of τ_c only if the effective potential of mean force acting on the two ends of the chain is known.

6. Appendix B

In formulating the fluctuations of the end-to-end distance vector, $\langle \delta \mathbf{R}_{ee}^2 \rangle$, it is important to take into account the failings of the continuum model of the freely jointed chain. A simple calculation of $\langle \delta \mathbf{R}_{ee}^2(t) \rangle$ with $\mathbf{R}_{ee}(t) = \mathbf{r}(N,t) - \mathbf{r}(0,t)$ as determined from eq 1 gives

$$\langle \delta \mathbf{R}_{ee}^2(t) \rangle = 16 N b^2 \sum_{n\ \mathrm{odd}} \frac{1}{n^2 \pi^2} (1 - \mathrm{e}^{-n^2 t/\tau_R}) \tag{29}$$

We will refer to this result as the standard analytic average. However, the nonphysical boundary conditions imposed on the continuum representation, with $\partial \mathbf{r}/\partial s \equiv 0$ at the endpoints, will strongly affect the accuracy of this result.

To minimize the effect of the boundary conditions on averages involving the end-to-end distance, we compute averages with respect to the differences between the centers of mass of the first and last bonds using

$$\mathbf{R}_{ee}(t) \approx \int_{N-1}^{N} \mathrm{d}s\, \mathbf{r}(s,t) - \int_0^1 \mathrm{d}s\, \mathbf{r}(s,t) \tag{30}$$

We will refer to this as the center of mass average. Using this representation, $\langle \delta \mathbf{R}_{ee}^2(t) \rangle$ is given in eq 4.

We fit the time-dependent diffusion coefficent (defined in eq 5), measured in simulations with $N = 19$ and $b = 0.39$, using both the standard analytic average (eq 29) and the center of mass average (eq 4), with the Kuhn length b taken as a fitting parameter for both average techniques. The results are shown in Figure 9. The center of mass average, which fits the data quite well, has a best fit of $b = 0.41$ (a difference of 5%), whereas the standard average does not give accurate results. For this reason, all averages involving $\mathbf{R}_{ee}$ are computed using the center of mass theory.

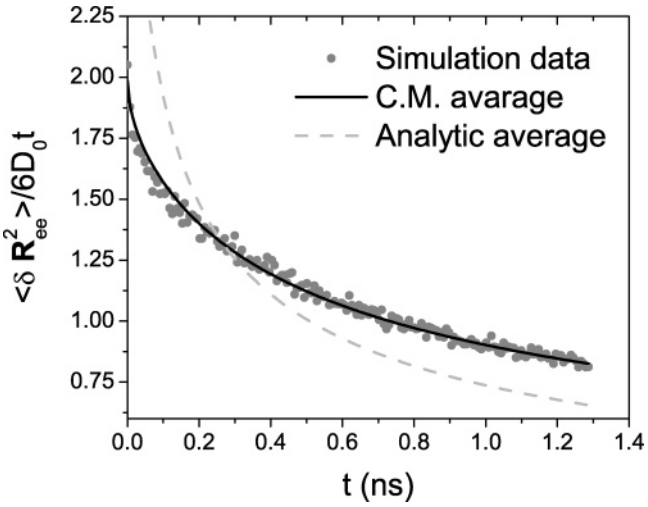

Figure 9. Measured diffusion coefficient as a function of time for the Rouse chain with $N = 19$ and $b = 0.39$ nm. Symbols are the simulation data, the dashed line (analytic average) is obtained using eqs 29 and 5 (with best fit of $b \approx 0.26$ nm), and the solid line is the center-of-mass average derived using eqs 4 and 5 (with best fit of $b \approx 0.41$ nm).

7. Appendix C

The relation between the mean first passage time τ and the probability $\Sigma(t)$ that at time t the system is still unreacted is exact

$$\tau = \int_0^\infty \Sigma(t) \mathrm{d}t \tag{26}$$

for any form of $\Sigma(t)$ for which $\Sigma(0)$ is finite and

$$\lim_{t \to \infty} t\Sigma(t) = 0$$

Therefore, the stricter requirement that $\Sigma(t) \sim \exp(-t/\tau)$ in the original SSS paper[4] is not required.

We define $F(t)$, the flux (or density) of passage, as $F(t) \equiv -\partial\Sigma(t)/\partial t$. The mean first passage time is

$$\tau = \int_0^\infty tF(t)\mathrm{d}t = \int_0^\infty t\left(-\frac{\partial\Sigma(t)}{\partial t}\right)\mathrm{d}t = -\int_0^\infty t\mathrm{d}\Sigma(t) \quad (27)$$

Performing integration by parts gives

$$\tau = -t\Sigma(t)|_0^\infty + \int_0^\infty \Sigma(t)\mathrm{d}t \quad (28)$$

By definition, $\Sigma(t)$ must be finite, and hence, $t\Sigma(t) = 0$ at $t = 0$. If $\Sigma(t)$ is such that it vanishes at $t \rightarrow \infty$ faster than t^{-1}, then the first term in eq 28 vanishes, and we are left with eq 26. Note that these are also necessary and sufficient conditions for τ in eq 26 to be finite.

References and Notes

(1) Winnik, M. A. In *Cyclic Polymers*; Semlyen, J. A., Ed.; Elsivier: New York, 1986; Chapter 9.

(2) Flory, P. J. *Principles of Polymer Chemistry*; Cornell University Press: Ithaca, NY, 1971.

(3) Wilemski, G.; Fixman, M. *J. Chem. Phys.* **1974**, *60*, 866−877.

(4) Szabo, A.; Schulten, K.; Schulten, Z. *J. Chem. Phys.* **1980**, *72*, 4350−4357.

(5) Doi, M. *Chem. Phys.* **1975**, *9*, 455−466.

(6) Sokolov, I. *Phys. Rev. Lett.* **2003**, *90*, 080601/1−080601/4.

(7) Debnath, P.; Cherayil, B. *J. Chem. Phys.* **2004**, *120*, 2482−2489.

(8) Guo, Z.; Thirumalai, D. *Biopolymers* **1995**, *36*, 83−102.

(9) Toan, N. M.; Marenduzzo, D.; Cook, P. R.; Micheletti, C. *Phys. Rev. Lett.* **2006**, *97*, 178302/1−178302/4.

(10) Cloutier, T. E.; Widom, J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 3645−3650.

(11) Du, Q.; Smith, C.; Shiffeldrim, N.; Vologodskaia, M.; Vologodskii, A. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 5397−5402.

(12) Lapidus, L.; Eaton, W.; Hofrichter, J. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 7220−7225.

(13) Lapidus, L.; Steinbach, P.; Eaton, W.; Szabo, A.; Hofrichter, J. *J. Phys. Chem. B.* **2002**, *106*, 11628−11640.

(14) Chang, I.-J.; Lee, J. C.; Winkler, J. R.; Gray, H. B. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 3838−3840.

(15) Lee, J. C.; Lai, B. T.; Kozak, J. J.; Gray, H. B.; Winkler, J. R. *J. Phys. Chem. B* **2007**, *111*, 2107−2112.

(16) Hudgins, R.; Huang, F.; Gramlich, G.; Nau, W. *J. Am. Chem. Soc.* **2002**, *124*, 556−564.

(17) Sahoo, H.; Roccatano, D.; Zacharias, M.; Nau, W. M. *J. Am. Chem. Soc.* **2006**, *128*, 8118−8119.

(18) Hagen, S. J.; Carswell, C. W.; Sjolander, E. M. *J. Mol. Biol.* **2001**, *305*, 1161−1171.

(19) Neuweiler, H.; Schulz, A.; Bohmer, M.; Enderlein, J.; Sauer, M. *J. Am. Chem. Soc.* **2003**, *125*, 5324−5330.

(20) Thirumalai, D.; Hyeon, C. *Biochemistry* **2005**, *44*, 4957−4970.

(21) Shore, D.; Baldwin, R. *J. Mol. Biol.* **1983**, *170*, 957−981.

(22) Bieri, O.; Wirz, J.; Hellrung, B.; Schutkowski, M.; Drewello, M.; Kiefhaber, T. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9597−9601.

(23) Woodside, M.; Behnke-Parks, W.; Larizadeh, K.; Travers, K.; Herschlag, D.; Block, S. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 6190.

(24) Woodside, M.; Anthony, P.; Behnke-Parks, W.; Larizadeh, K.; Herschlag, D.; Block, S. *Science* **2006**, *314*, 1001−1004.

(25) Portman, J. *J. Chem. Phys.* **2003**, *118*, 2381−2391.

(26) Camacho, C.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 1277−1281.

(27) Socci, N. D.; Onuchic, J. N.; Wolynes, P. G. *J. Chem. Phys.* **1996**, *104*, 5860−5868.

(28) Pastor, R.; Zwanzig, R.; Szabo, A. *J. Chem. Phys.* **1996**, *105*, 3878−3882.

(29) Sakata, M.; Doi, M. *Polym. J.* **1976**, *8*, 409−413.

(30) Chen, J.; Tsao, H.-K.; Sheng, Y.-J. *Phys. Rev. E* **2005**, *72*, 031804/1−031804/7.

(31) Podtelezhnikov, A.; Vologodskii, A. *Macromolecules* **1997**, *30*, 6668−6673.

(32) de Gennes, P. G. *J. Chem. Phys.* **1982**, *76*, 3316−3321.

(33) Hyeon, C.; Thirumalai, D. *J. Chem. Phys.* **2006**, *124*, 104905/1−104905/14.

(34) Doi, M.; Edwards, S. *The Theory of Polymer Dynamics*; Oxford University Press: Oxford, U.K., 1986.

(35) Buscaglia, M.; Lapidus, L.; Eaton, W.; Hofrichter, J. *Biophys. J.* **2006**, *91*, 276−288.

(36) Hyeon, C.; Dima, R.; Thirumalai, D. *Structure* **2006**, *14*, 1633−1645.

(37) Frenkel, D.; Smit, B. *Understanding Molecular Simulation—From Algorithms to Applications*; Academic Press: San Diego, CA, 2002.

(38) Rissanou, A.; Anastasiadis, S.; Bitsanis, I. *J. Polym. Sci., Part B: Polym. Phys.* **2006**, *44*, 3651−3666.

(39) Thirumalai, D. *J. Phys. Chem. B* **1999**, *103*, 608−610.

(40) Jun, S.; Bechhoefer, J.; Ha, B.-Y. *Europhys. Lett.* **2003**, *64*, 420−426.

(41) Laurence, T. A.; Kong, X.; Jäger, M.; Weiss, S. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 17348−17353.

(42) Hoffmann, A.; Kane, A.; Nettels, D.; Hertzog, D. E.; Baumgärtel, P.; Lengefeld, J.; Reichardt, G.; Horsley, D. A.; Seckler, R.; Bakajin, O.; Schuler, B. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 105−110.

(43) Friedman, B.; O'Shaughnessy, B. *J. Phys. II* **1991**, *1*, 471−486.

(44) de Gennes, P. G. *J. Chem. Phys.* **1982**, *76*, 3316−3321.

(45) de Gennes, P. G. *Scaling Concepts in Polymer Physics*; Cornell University Press: Ithaca, NY, 1979.

(46) des Cloizeaux, J. *J. Phys.* **1980**, *41*, 223−238.

(47) Ortiz-Repiso, M.; Rey, A. *Macromolecules* **1998**, *31*, 8363−8369.