Kinetics of Interior Loop Formation in Semiflexible Chains

Changbong Hyeon and D. Thirumalai

Biophysics Physics Program,

Institute for Physical Science and Technology

University of Maryland,

College Park, MD 20742

Loop formation between monomers in the interior of semiflexible chains describes elementary events in biomolecular folding and DNA bending. We calculate analytically the interior distance distribution function for semiflexible chains using a mean-field approach. Using the potential of mean force derived from the distance distribution function we present a simple expression for the kinetics of interior looping by adopting Kramers theory. For the parameters, that are appropriate for DNA, the theoretical predictions in comparison to the case are in excellent agreement with explicit Brownian dynamics simulations of worm-like chain (WLC) model. The interior looping times (τ_{IC}) can be greatly altered in cases when the stiffness of the loop differs from that of the dangling ends. If the dangling end is stiffer than the loop then τ_{IC} increases for the case of the WLC with uniform persistence length. In contrast, attachment of flexible dangling ends enhances rate of interior loop formation. The theory also shows that if the monomers are charged and interact via screened Coulomb potential then both the cyclization (τ_c) and interior looping (τ_{IC}) times greatly increase at low ionic concentration. Because both τ_c and τ_{IC} are determined essentially by the effective persistence length ($l_p^{(R)}$) we computed $l_p^{(R)}$ by varying the range of the repulsive interaction between the monomers. For short range interactions $l_p^{(R)}$ nearly coincides with the bare persistence length which is determined largely by the backbone chain connectivity. This finding rationalizes the efficacy of describing a number of experimental observations (response of biopolymers to force and cyclization kinetics) in biomolecules using WLC model with an effective persistence length.

I. INTRODUCTION

The kinetics of formation of contact between the ends of a polymer chain has a rich history.[1,2] Both experiments,[1,2] theory,[3,4,5,6,7,8,9] and simulations[10,11,12] have been used to address the elementary event of the dynamics of end-to-end contact formation (or cyclization kinetics) (Fig.1-A). Contact formation between two reactive groups separated by a certain distance along the chain is a basic intramolecular rate process in a polymer. Recently, there has been renewed interest in understanding the looping dynamics that has been studied both theoretically[3,4,5,6,7] and experimentally[12,13,14,15,16] because of its fundamental importance in a number of biological processes. The hairpin loop formation is the elementary step in RNA folding,[17] structure formation in ssDNA,[18,19] and protein folding.[13,20,21,22,23] Cyclization in DNA has recently drawn renewed attention not only because of its importance in gene expression[24,25] but also it provides a way to assess DNA's flexibility. The promise of using single molecule technique to probe the real time dynamics of polymer chains has also spurred theories and simulations of cyclization kinetics. Using loop formation times between residues that are in the interior as the most elementary event in protein folding, it has been argued, using experimental data and theoretical expression for probability for loop formation in stiff chains, that the speed limit for folding is on the order of a 1 μs.[26] These examples illustrate the need to understand quantitatively the elementary event of contact formation between segments of a polymer chain.

Even without taking hydrodynamic interactions into account theoretical treatment of cyclization kinetics in polymer chains is difficult because several relaxation times and length and energy scales are interwined. At the minimum the variation of time scale for cyclization (τ_c) with polymer length is dependent on polymer relaxation time (τ_R). In biopolymers additional considerations due to chain stiffness and heterogeneity of interactions between monomer (amino acid residue or nucleotides) must be also taken into account. Majority of the cyclization kinetics studies on synthetic polymers[2] have considered examples in which the contour length (L) of the polymer is much greater than its persistence length (l_p). In contrast, loop formation dynamics in biopolymers have focused on systems in which L/l_p is relatively small. In disordered polypeptide chains L/l_p can be as small as 3,[15,27] while in DNA $L/l_p < 1$.[16,28] Thus, it is important to develop theoretical tools for the difficult problem of loop formation dynamics for arbitrary L

and l_p. Despite the inherent complexities in treating loop formation in biopolymers it has been found that the use of polymer-based approach is reasonable in analyzing experimental data on cyclization kinetics in proteins[8,15] and DNA.[16]

In this paper we are primarily concerned with the looping dynamics between interior segments of a semiflexible chain. While a lot of theoretical and experimental works (mentioned above) have been done on the end-to-end looping (Fig.1-(a)), only a few studies have been reported on the contact formation between monomers in the interior of a chain (interior looping) (Fig.1-(b)).[29,30,31,32,33] There are a few reasons to consider kinetics of interior looping. (1) The biological events such as hairpin formation and DNA looping often involve contact formation between monomers that are not at the ends of the chain. For example, it is thought that the initiation of nucleation in protein folding occurs at residues that are near the loop regions.[34] The residues that connect these loops are in the interior of the polypeptide chain. Similar processes are also relevant in RNA folding.[35] (2) It is known that for flexible chains with excluded volume interactions (polymer in a good solvent) the probability of loop formation is strongly dependent on the location of the two segments. For large loop length (S) the loop formation probability, $P(S)$, in three dimensions for chain ends $\sim S^{\theta_1}$ where $\theta_1 \approx 1.9$ while $P(S) \sim S^{\theta_2}$ with $\theta_2 \sim 2.1$ for monomer in the interior.[36] Although the values of θ_1 and θ_2 are similar it could lead to measurable differences in loop formation times.[32]

The rest of the paper is organized as follows. In section II we present the physical considerations that give rise to the well-established results for τ_c for flexible chains. The extension of the arguments for flexible chains to semiflexible polymers suggests that the local equilibrium approximation can be profitably used to analyze both cyclization kinetics and interior looping dynamics. The basic theory for the equilibrium distance distribution between two interior segments s_1 and s_2 (Fig.1-(b)) is presented in section III. Using the equilibrium distribution function and adopting Kramers theory and following the suggestion by Jun *et. al.*,[37] we obtain an analytical expression for time scale τ_{IC} for interior contact formation in section IV. Explicit results of simulations of worm-like chain (WLC), which validate the theory, are presented in section V. In section VI we consider the kinetics of interior loop formation in WLC in which the stiffness of the loop is different from that of the dangling ends. Section VII describes the consequences of screened Coulomb interaction between monomer segments on cyclization

kinetics and interior looping dynamics. Because the results in section VIII are expressed in terms of a renormalized persistence length $(l_p^{(R)})$ of WLC we present simulation results for $l_p^{(R)}$ variation for a number of potentials that describe interactions between monomers in section VII. The conclusions of the article are summarized in section IX.

II.   PRELIMINARY CONSIDERATIONS

The pioneering treatment of loop formation dynamics due to Wilemski and Fixman (WF)[3,38] has formed the basis for treating cyclization kinetics in flexible polymer chains. Using a generalized diffusion equation for the probability density, $\frac{\partial P(\{\mathbf{r}^N\},t)}{\partial t} = \mathcal{L}_{FP}P(\{\mathbf{r}^N\},t) - k\mathcal{S}(\{\mathbf{r}^N\})P(\{\mathbf{r}^N\},t)$ ($\mathcal{L}_{FP}$ is a generalized diffusion operator, $k\mathcal{S}$ is a sink term) for a N-segment polymer, and local equilibrium approximation within the sink, WF expressed the cyclization time τ_c in terms of an integral involving a sink-sink correlation function. From the WF formalism and related studies it is known that even in the simplest cases (ideal chains or polymers with excluded volume interactions) the validity of the local equilibrium approximation depends on the interplay between τ_c and the chain relaxation time, τ_R. If $\tau_c \gg \tau_R$ then the local equilibrium approximation is expected to hold because the polymer chain effectively explores the available volume *before* the monomers at the end (reactive groups) form a contact. In this situation, τ_c can be computed by considering mutual diffusion of the chain ends in a potential of mean force $(F(R_e))$. For ideal chains, $F(R_e) = -k_B T \log P(R_e) \sim 3k_B T R_e^2/2\overline{R}^2$ where R_e is the end-to-end distance, $\overline{R} \sim aN^{1/2}$ is the mean end-to-end distance, a is the size of the monomer, T is the temperature, and k_B is the Boltzmann constant. By solving such an equation subject to the absorbing boundary condition, Szabo, Schulten, and Schulten (SSS)[6] showed that $\tau_{SSS} = \tau_o N^{3/2}$. Simulations[4] and theory[5] show that if the capture radius for contact formation is non-zero, and is on the order of a monomer size then $\tau_c \sim \frac{\langle R^2 \rangle}{D_c} \sim \tau_1 N^{2\nu+1}$ where $\nu = 1/2$ for Rouse chains and $\nu \approx 3/5$ for polymers with excluded volume, and D_c is a mutual diffusion coefficient. The use of these theories to analyze the dependence of τ_c on N in polypeptides shows that the physics of cyclization kinetics is reasonably well described by diffusion in a potential of mean force $F(R_e)$ which only requires accurate calculation of $P(R_e)$ the end-to-end distri-

bution function.[15,37] For describing interior looping times τ_{IC} for contact between two interior monomers s_1 and s_2 we need to compute $P(R_{12}, |s_1 - s_2|)$ where R_{12} is the distance between s_1 and s_2. With $P(R_{12}, |s_2 - s_1|)$ in hand τ_c can be computed by solving a suitable diffusion equation.

Because the use of $F(R_e)$ in computing τ_c and τ_{IC} is intimately related to chain relaxation times it is useful to survey the conditions which satisfy the local equilibrium approximation. By comparing the conformational space explored by the chain ends compared to the available volume prior to cyclization[39] the validity of the local equilibrium approximation in flexible chains can be expressed in terms of an exponent $\theta = \frac{d+g}{z}$.[7] Here d is the spatial dimension, the correlation hole exponent (des Cloizeaux exponent)[40] g describes the probability of the chain ends coming close together, and z is the dynamical scaling exponent ($\tau_R \sim \overline{R}^z$). If $\theta > 1$ the local equilibrium approximation is expected to hold and τ_c is determined essentially by the equilibrium $P(R_e)$ as $R_e \to a$ the capture radius. Using the scaling form of $P(R_e)$ for small R_e $P(R_e) \sim \frac{1}{\overline{R}^d}\left(\frac{R_e}{\overline{R}}\right)^g$ and $\overline{R} \sim N^\nu$ (ν is the Flory exponent) we find $\tau_c \sim N^{\nu(d+g)}$. For Gaussian chains $\nu = 1/2$ and $g = 0$ and hence $\tau_{SSS} \sim \tau_c \sim N^{3/2}$. This result was obtained fifty years ago by Jacobsen and Stockmayer.[41] However, in the free-draining case ($z = 4$, $g = 0$, $\nu = 1/2$, $d = 3$), $\theta < 1$ and hence the condition $\tau_c \gg \tau_R$ is not satisfied. In this case $\tau_c \sim \tau_R \sim N^{z\nu} \sim N^2$. Thus, for ideal Gaussian chains it is likely that $\tau_{SSS} < \tau_c < \tau_{WF}$.[42] Indeed, recent simulations show that if the number of statistical segments is large ($\gtrsim 20$) then for ideal chains $\tau_c \sim N^2$ which signals the breakdown of the condition $\tau_c \gg \tau_R$. Experiments on cyclization of polypeptide chains show that $\tau_c \sim N^{3/2}a$ is obeyed for N in the range $10 < N < 20$ (see Fig.(5) in Ref.[15]). Deviations from ideal chain results are found for $N < 10$, either due to chain stiffness[15] or sequence variations.[43] For polymer chains in good solvents with hydrodynamic interactions ($d = 3$, $g = 5/18$, $z = 3$, and $\nu = 3/5$), $\theta = 59/54 > 1$. Thus, in real chains the local equilibrium approximation may be accurate.

For stiff chains bending rigidity severely restricts the allowed conformations especially when the contour length (L) is on the order of the persistence length (l_p). Because of high bending rigidity the available volume is restricted by thermal fluctuations. Clearly in this situation, the chain is close to equilibrium. This may be the case for short DNA segments. In effect these chains satisfy the $\tau_c > \tau_R$ condition which enables us to calculate τ_c or τ_{IC} by solving an

appropriate one dimensional diffusion equation (see below) in a suitable potential of mean force.

Effect of chain Stiffness : Many biopolymers are intrinsically stiff and are better described by worm-like chain (WLC) models. The persistence length, which is a measure of stiffness, varies considerably. It ranges from $(3-7)$ Å (proteins),[15] $(10-25)$ Å (ss-DNA[19,44] and RNA,[45,46]) 50 nm for ds-DNA. Typically, loops of only a few persistence length form, which underscores the importance of chain stiffness. In order to correctly estimate the loop closure time, consideration of the stiffness in the loop closure dynamics is necessary unless the polymer looping takes place between the reactive groups that are well separated and the chain length L is long. If $L \gg l_p$ (persistence length), the looping dynamics will follow the scaling law for flexible chains. However, at short length scales loop dynamics can be dominated by chain stiffness.[16] If the chain is stiff then WLC conformations are limited to those allowed by thermal fluctuations. In this situation, the time for exploring the chain conformations is expected to be less than τ_c. Thus, we expect local equilibrium to be a better approximation for WLC than for long flexible chains.

Recently, Dua *et. al.*[47] have studied the effect of stiffness on the polymer dynamics based on Wilemski-Fixman formalism and showed, that for free-draining semiflexible chain without excluded volume $\tau_c \sim N^{2.2 \sim 2.4}$ at moderate values of stiffness. However, the procedure used to obtain this result is not complete, as recognized by the authors, because they use a Gaussian propagator $\mathcal{G}(\mathbf{r}, t | \mathbf{r}', 0) = \left(\frac{3}{2\pi \langle \mathbf{r}^2 \rangle (1 - \phi(t))} \right)^{3/2} \exp \left(-\frac{3(\mathbf{r} - \phi(t)\mathbf{r}')^2}{2\langle \mathbf{r}^2 \rangle (1 - \phi(t))} \right)$ which is not valid for WLC. The end-to-end distance distribution becomes a Gaussian at equilibrium, $\lim_{t \to \infty} \mathcal{G}(\mathbf{r}, t | \mathbf{r}', 0) = P_{eq}(\mathbf{r}) = \left(\frac{3}{2\pi \langle \mathbf{r}^2 \rangle} \right)^{3/2} \exp \left(-\frac{3\mathbf{r}^2}{2\langle \mathbf{r}^2 \rangle} \right)$, which is incorrect for semiflexible chain especially when $l_p \sim L$ (see Fig.2 and Refs.[48,49]).

As an alternative method we include the effect of chain stiffness assuming that local equilibrium approximation is valid. This is tantamount to assuming that $\tau_c > \tau_R$ which, for reasons given above, may be an excellent approximation for WLC.[37] In this case we can compute τ_c by solving the diffusion equation in a one dimensional potential $F(R_e) = -k_B T \ln P(R_e)$ where $P(R_e)$ is the probability of end-to-end distance distribution for WLC. For the problem of interest, namely, the computation of τ_{IC}, we generalize the approach of Jun *et. al.*[37] who used Kramers theory in the effective potential $F(R_e)$ to obtain τ_c. In general, the time for cyclization

can be calculated using

$$\tau_c = \int_a^r dy e^{\beta F(y)} \frac{1}{D} \int_y^L dz e^{-\beta F(z)} \tag{1}$$

where a is the capture (contact) radius of the two reactive groups. We show that Eq.(1) provides accurate estimates of τ_c, thus suggesting the local equilibrium approximation is guaranteed.

Here, we address the following specific questions: What is the loop formation time between the interior segments in a semiflexible chain? Does the dangling ends (Fig.1-(b)) affect the dynamics of loop formation? How does the effect of interaction between monomer segments (e.g. excluded volume, electrostatic interaction) affect loop closure kinetics in WLC models?

III.   DISTANCE DISTRIBUTION FUNCTION BETWEEN TWO INTERIOR POINTS

A key ingredient in the calculation of the potential of mean force is appropriate distribution function between the two monomers that form a contact. In Refs.[49,50] the equilibrium end-to-end (R_e) radial distribution function of a semiflexible chain $P(R_e)$ was obtained in terms of the persistence length (l_p) and the contour length (L). Despite the mean field approximation employed in Refs.[49,50] the distribution function $P(R_e)$ is in very good agreement with simulations.[48] The simplicity of the final expression has served as a basis for analyzing a number of experiments on proteins,[51] RNA[45] and DNA.[52] In this section, we use the same procedure to calculate the distribution function $P(R_{12}; l_p, s_1, s_2, L)$ where $0 < s_1, s_2 < L$, and R_{12} is the spatial distance between s_1 and s_2.

For the semiflexible chain in equilibrium we write the distribution function of the distance R_{12} between s_1 and s_2 (Fig.1) along the chain contour as

$$\begin{aligned}
G(\mathbf{R}_{12}; s_1, s_2) &= \langle \delta(\mathbf{R}_{12} - \int_{s_1}^{s_2} \mathbf{u}(s) ds) \rangle_{MF} \\
&= \frac{\int D[\mathbf{u}(s)] \delta(\mathbf{R}_{12} - \int_{s_1}^{s_2} \mathbf{u}(s) ds) \Psi_{MF}[\mathbf{u}(s)]}{\int D[\mathbf{u}(s)] \Psi_{MF}[\mathbf{u}(s)]}
\end{aligned} \tag{2}$$

where $\mathbf{u}(s)$ is a unit tangent vector at position s. The exact weight for the semiflexible chain is $\Psi[\mathbf{u}(s)] \propto \exp\left[-\frac{l_p}{2} \int_0^L ds \left(\frac{\partial \mathbf{u}}{\partial s}\right)^2\right] \prod \delta(\mathbf{u}^2(s) - 1)$. The nonlinearity, that arises due to the restriction $\mathbf{u}^2(s) = 1$, makes the computation of the path integral in Eq.(2) difficult. To circumvent

the problem we replace $\Psi[\mathbf{u}(s)]$ by the mean field weight $\Psi_{MF}[\mathbf{u}(s)]$,[53]

$$\Psi_{MF}[\mathbf{u}(s)] \propto \exp\left[-\frac{l_p}{2}\int_0^L \left(\frac{\partial \mathbf{u}(s)}{\partial s}\right)^2 ds - \lambda \int_0^L (\mathbf{u}^2(s) - 1)ds - \delta[(\mathbf{u}_0^2 - 1) + (\mathbf{u}_L^2 - 1)]\right]. \quad (3)$$

The Lagrange multipliers λ and δ, which are used to enforce the constraint $\mathbf{u}^2(s) = 1$,[54] will be determined using stationary phase approximation (see below). The path integral associated with the weight $\Psi_{MF}[\mathbf{u}(s)]$ is equivalent to a kicked quantum mechanical harmonic oscillator with "mass" l_p and angular frequency $\Omega \equiv \sqrt{2\lambda/l_p}$. Using the propagator for the harmonic oscillator

$$Z(\mathbf{u}_s, \mathbf{u}_0, s) = \left(\frac{\pi \sinh(\Omega s)}{\Omega_p}\right)^{-\frac{3}{2}} \exp\left(-\Omega_p \frac{(\mathbf{u}_s^2 + \mathbf{u}_0^2)\cosh(\Omega s) - 2\mathbf{u}_s \cdot \mathbf{u}_0}{\sinh(\Omega s)}\right) \quad (4)$$

and defining $\Omega_p \equiv \frac{\Omega l_p}{2}$ the isotropic distribution function becomes

$$\begin{aligned}
G(R_{12}, s_1, s_2) &= \mathcal{N}^{-1} e^{\lambda L + 2\delta} \int \frac{d^3\mathbf{k}}{(2\pi)^3} \int d\mathbf{u}_0 d\mathbf{u}_{s_1} d\mathbf{u}_{s_2} d\mathbf{u}_L e^{-\delta \mathbf{u}_0^2} Z(\mathbf{u}_0, \mathbf{u}_{s_1}; s_1) \\
&\times e^{i\mathbf{k}\cdot\mathbf{R}_{12} - \frac{\mathbf{k}^2}{4\lambda}|s_1 - s_2|} Z\left(\mathbf{u}_{s_1} + \frac{i\mathbf{k}}{2\lambda}, \mathbf{u}_{s_2} + \frac{i\mathbf{k}}{2\lambda}; s_2 - s_1\right) \\
&\times e^{-\delta \mathbf{u}_L^2} Z(\mathbf{u}_{s_2}, \mathbf{u}_L; L - s_2).
\end{aligned} \quad (5)$$

By writing the distribution function as $G(R_{12}, s_1, s_2) = \int_{-i\infty}^{i\infty} d\lambda \int_{-i\infty}^{i\infty} d\delta \exp(-\mathcal{F}[\lambda, \delta])$ it is clear that the major contribution to G (in the thermodynamic limit $L \to \infty$) comes from the saddle points of the free energy functional $\mathcal{F}[\lambda, \delta]$, i.e., $\frac{\partial \mathcal{F}}{\partial \lambda} = \frac{\partial \mathcal{F}}{\partial \delta} = 0$. The functional $\mathcal{F}[\lambda, \delta]$ is (see Appendix A for details of the derivation)

$$\begin{aligned}
\mathcal{F}[\lambda, \delta] &= -(L\lambda + 2\delta) \\
&+ \frac{3}{2}\ln\left(\frac{\sinh \Omega L}{\Omega_p}(\delta^2 + \Omega_p^2 + 2\delta\Omega_p \coth \Omega L)\right) - \frac{3}{2}\ln\frac{\lambda^2}{Q(s_1, s_2; \lambda, \delta)} \\
&+ \frac{\lambda^2 R_{12}^2}{Q(s_1, s_2; \lambda, \delta)}
\end{aligned} \quad (6)$$

To obtain the optimal values of λ and δ we first take the $L \to \infty$ limit and then solve stationarity conditions $\frac{\partial \mathcal{F}}{\partial \lambda} = \frac{\partial \mathcal{F}}{\partial \delta} = 0$. Technically, the optimal value of δ and λ should be calculated for a given L and then it is proper to examine the $L \to \infty$ limit. The consequences of reversing the order of operation are discussed in Appendix B. Using the first procedure (taking $L \to \infty$ first) we obtain $Q(s_1, s_2; \lambda, \delta) \to |s_2 - s_1|\lambda$ in the limit $L \gg s_2 \gg s_1 \gg 1$, and thus

$\mathcal{F}[\lambda, \delta]$ becomes

$$
\begin{aligned}
\mathcal{F}[\lambda, \delta] \approx\ & -(L\lambda + 2\delta) \\
& + \frac{3}{2}\ln[\frac{e^{\Omega L}}{\Omega_p}(\frac{\delta}{\Omega_p} + 1)^2] + \frac{3}{2}\ln\Omega_p^2 + \frac{3}{2}\ln\frac{|s_2 - s_1|}{\lambda} + \frac{\mathbf{R}_{12}^2\lambda}{|s_2 - s_1|} \\
=\ & L\left(\frac{3}{2}\Omega - \lambda(1 - \frac{|s_2 - s_1|}{L}\frac{\mathbf{R}_{12}^2}{|s_2 - s_1|^2})\right) \\
& + \frac{3}{2}\ln[\frac{\Omega_p}{\lambda}(\frac{\delta}{\Omega_p} + 1)^2] + \frac{3}{2}\ln|s_2 - s_1| - 2\delta
\end{aligned} \tag{7}
$$

where we have omitted numerical constants. The major contribution to the integral over λ and δ comes from the sets of λ and δ which pass the saddle point of a stationary phase contour on the $\mathbf{Re}\{\mathcal{F}\}$ plane. Since the term linear in L dominates the logarithmic term in l_p even when $L/l_p \sim \mathcal{O}(1)$, the stationary condition for λ can be found by taking the derivative with respect to λ by considering only the leading term in L (cf. see Appendix B for details). The stationarity condition leads to

$$
\Omega_p = \sqrt{\frac{\lambda l_p}{2}} = \frac{3}{4}\frac{1}{1 - \frac{|s_2 - s_1|}{L}r^2} \tag{8}
$$

where $\mathbf{r} = \frac{\mathbf{R}_{12}}{|s_2 - s_1|}$ with $0 < r < 1$. Similarly, the condition for δ can be obtained as

$$
\delta = \frac{3}{2} - \Omega_p. \tag{9}
$$

Determination of the parameters λ and δ by the stationary phase approximation amounts to replacing the local constraint $\mathbf{u}^2(s) = 1$ by a global constraint $\langle\mathbf{u}^2(s)\rangle = 1$.[54] Finally, the stationary values of λ and δ in the large L limit give the interior distance distribution function:

$$
G(R_{12}, s_2 - s_1) = \frac{N}{(1 - \frac{|s_2 - s_1|}{L}r^2)^{9/2}}\exp\left(-\frac{9|s_2 - s_1|}{8l_p(\frac{|s_2 - s_1|}{L})(1 - \frac{|s_2 - s_1|}{L}r^2)}\right). \tag{10}
$$

The mean-field approximation allows us to obtain a simple expression for the internal segment distance distribution function. The previously computed $P(R_e)$[49] can be retrieved by setting $|s_2 - s_1| = L$. The radial probability density, for the interior segments, in three dimensions, for semiflexible chains is

$$
P(r; s_2 - s_1, t) = 4\pi C\frac{r^2}{(1 - \frac{|s_2 - s_1|}{L}r^2)^{9/2}}\exp\left(-\frac{3t}{4(\frac{|s_2 - s_1|}{L})(1 - \frac{|s_2 - s_1|}{L}r^2)}\right) \tag{11}
$$

where $r = R_{12}/|s_2 - s_1| \equiv R/|s_2 - s_1|$ and $t = |s_2 - s_1|/l_0$ with $l_0 = \frac{2}{3}l_p$. The normalization constant C is determined using $\int_0^1 P(r, s; t)dr = 1$. The integral is evaluated by the substitution $\sqrt{\frac{|s_2 - s_1|}{L}}r = \frac{x}{\sqrt{1+x^2}}$ to yield

$$
\begin{aligned}
C &= \frac{1}{4\pi}(\frac{|s_2 - s_1|}{L})^{3/2}\left(\int_0^{x_0} dx x^2(1 + x^2)e^{-\alpha(1+x^2)}\right)^{-1} \\
&= \frac{4}{\pi\alpha^{-7/2}}(\frac{|s_2 - s_1|}{L})^{3/2}[-2\sqrt{\alpha}x_0 e^{-\alpha(1+x_0^2)}(15 + 2\alpha(6 + 5x_0^2 + 2\alpha(1 + x_0^2)^2)) \\
&+ \alpha^2 e^{-\alpha}\sqrt{\pi}\text{erf}[\sqrt{\alpha}x_0](1 + 3\alpha^{-1} + \frac{15}{4}\alpha^{-2})]^{-1}
\end{aligned}
\tag{12}
$$

where $\alpha = \frac{3t}{4\frac{|s_2 - s_1|}{L}}$, $x_0 = \sqrt{\frac{|s_2 - s_1|}{L - |s_2 - s_1|}}$, and $\text{erf}(x)$ is the error function. The peak in the distribution function is at

$$
r_{max} = \sqrt{\frac{\eta + \sqrt{\eta^2 + 14}}{7|s_2 - s_1|/L}}
\tag{13}
$$

where $\eta = \frac{5}{2} - \frac{3t}{4\frac{|s_2 - s_1|}{L}}$. For $|s_2 - s_1| = L$, $r_{max} \to 0$ as $t \to \infty$ and $r_{max} \to 1$ as $t \to 0$.

In Fig.2 we compare the distribution functions $P(R_e)$ and $P(R_{12}, s_2 - s_1)$. When $|s_1 - s_2|/L = 1$ Eq.(11) gives the end-to-end distribution for semiflexible chains. By adjusting the value of t (or equivalently l_0) we can go from flexible to intrinsically stiff chains. As the chain gets stiff there is a dramatic difference between the $P(r; |s_1 - s_2|, t)$ and $P(r; |s_1 - s_2| = L, t)$ (see Fig.2-(b)). Contact formation between interior segments are much less probable than cyclization process (compare the green and red curves with the black in Fig.2-(b)). Physically, this is because stiffness on shorter length scales ($|s_1 - s_2|/L < 1$) is more severe than when $|s_1 - s_2|/L \sim \mathcal{O}(1)$. However, when the chain is flexible (large t) the difference between the probability of contact between the interior segments and cyclization is small (Fig.2-(a)). In the limit of large $t(\propto L/l_p)$ the Hamiltonian in Eq.(3) describes a Gaussian chain for which the distance distribution between interior points remains a Gaussian. However, if excluded volume interactions are taken into account there can be substantial difference between $P(x, |s_1 - s_2|, t)$ and end-to-end segment distribution even when t is moderately large.

IV.   INTERIOR LOOP CLOSURE TIME USING KRAMERS THEORY

Having obtained the effective potential between interior segments of semiflexible chain we can evaluate Eq.(1) using $F(R) = -k_B T \log P(R)$ with $P(R)$ given by Eq.(11) with $r = R/L$. For clarity we have suppressed the dependence of $P(R)$ on $|s_2 - s_1|$. The expression for the mean first passage time (Eq.(1)) can be approximated by expanding the effective potential $F(R)$ at the barrier top and at the the bottom as $F(R) \approx F(R_t) - \frac{1}{2}\nabla_R^2 F(R)|_{R=R_t}(R - R_t)^2 + \cdots$ and $F(R) \approx F(R_b) + \frac{1}{2}\nabla_R^2 F(R)|_{R=R_b}(R - R_b)^2 + \cdots$, respectively (Fig.2). Evaluating the resulting Gaussian integrals yields the Kramers equation

$$\tau_c \sim \tau_{Kr} = \frac{\pi k_B T}{D\sqrt{\nabla_R^2 F(R)|_{R=R_b}}\sqrt{\nabla_R^2 F(R)|_{R=R_t}}} \exp\left(\Delta F^{\ddagger}/k_B T\right). \tag{14}$$

When evaluating the Gaussian integral at the barrier top with $R = R_t$ (Fig.2), we assume that only the integral beyond $R > R_t$ contributes to the result. In the overdamped limit the mean first passage time, which is roughly the inverse of the reaction rate, is determined by the barrier height $(\Delta F^{\ddagger} = F(R_t) - F(R_b))$, and the curvatures of the bound state, the curvature at the barrier top, and the friction coefficient, that depends on $D(= 2D_0)$ where D_0 is the monomer diffusion coefficient. The curvatures of the potential at the bottom $(R_\alpha = R_b)$ or at the top $(R_\alpha = R_t)$ (Fig.2 right panel) is obtained using $\sqrt{\nabla_R^2 F(R)|_{R=R_\alpha}} = \sqrt{k_B T \left(\frac{6}{R_\alpha^2} - \frac{G''(R_\alpha,S,L)}{G(R_\alpha,S,L)}\right)}$ by imposing the condition $\nabla_R F(R)|_{R=R_\alpha} = 0$, i.e., $\left(\frac{2}{R} + \frac{\partial_R G(R,S,L)}{G(R,S,L)}\right)|_{R=R_\alpha} = 0$. There is an uncertainty in the evaluation $\sqrt{\nabla_R^2 F(R)|_{R=R_t}}$ because $F(R)$ does not really form a barrier at $R = R_t$. Thus, we assume that the curvature at the barrier top is $\sim 1/R_t$ using dimensional analysis. We express lengths in terms of the persistence length $l_p^{(0)} = l_0$. Setting $\frac{|s_1-s_2|}{l_p^{(0)}} \equiv s$, $\frac{L}{l_p^{(0)}} \equiv l$, and $\frac{R}{l_p^{(0)}} \equiv x$ the radial probability density is

$$P(x,s,l) = \frac{4\pi C(s,l)(\frac{x^2}{s^3})}{(1 - \frac{x^2}{ls})^{9/2}} \exp\left[-\frac{3s}{4(s/l)(1 - \frac{x^2}{ls})}\right]. \tag{15}$$

with $\int_0^s dx P(x,s,l) = 1$. When the dimensionless contact radius $x_t \equiv \alpha(= a/l_p^{(0)}) \ll 1$, the exponential factor can be approximated as $\exp\left(-\Delta F^{\ddagger}/k_B T\right) = P(\alpha,s,l)/P(x_b,s,l) \simeq \frac{\alpha^2 G(0,s,l)}{x_b^2 G(x_b,s,l)}$.

The function $P(r)$ (Eq.(11)) is not appropriate for estimating the contact probability of semiflexible chains even though the overall shape of the mean field distribution function is in excellent agreement with the simulations and experiment. The contact probability for DNA is

well studied by Shimada and Yamakawa,[55] thus we use their result for $G(0, l)$. If $x = 0$ and $s < 10$ then the Shimada-Yamakawa equation gives a reliable estimate of the *looping probability* $(G(0, s, l) = G_0(s))$.

$$G_0(s) = \frac{896.32}{s^5} \exp\left(-14.054/s + 0.246s\right) \tag{16}$$

At a large $s(> 10)$ value an interpolation formula $G_0(s) \sim s^{-3/2}$ due to Ringrose *et. al.*[56] can be used. Note that as the chain gets stiffer (l decrease) $G_0(s)$ decreases substantially indicating a great reduction in the loop formation probability for intrinsically stiff chains. It should be stressed that Eq.(16) has been obtained only for cyclization process and does not take into account the effect of dangling ends (Fig.1-(b)). The contact probability ($r \to 0$) between interior segments should be different from the one for the end-to-end contact. In other words $G(r \to 0, s, l)$ should depend on s/l. Unfortunately, we do not know of analytical results for $G(0, s, l)$. We simply use $G_0(s)$ for $G(0, s, l)$ and resort to the values of x_b and $G(x_b, s, l)$ to account for end effects. We validate the approximation that $G(0, s, l)$ does not depend on l $(G_0(s) = G(0, s, l))$ explicitly using simulations (see below).

With these approximations the loop formation time in the presence of dangling ends (Fig.1-(b)) is

$$\tau_c(s, l) \simeq \left\{\frac{\pi}{\xi} x_b^2 G(x_b, s, l) / \left(\frac{6}{x_b^2} - \frac{G''(x_b, s, l)}{G(x_b, s, l)}\right)^{1/2}\right\} \frac{1}{\alpha D} \frac{(l_p^{(0)})^2}{G_0(s)}. \tag{17}$$

where $x = x_b = \sqrt{\frac{\eta + \sqrt{\eta^2 + 14}}{7}} sl$ and ξ is the adjustable parameter we introduced to account for the uncertainty in computing the curvature at the barrier top, i.e., $\sqrt{\nabla_x^2 F(x)}|_{x=x_t \equiv \alpha} = \frac{\xi}{\alpha}\sqrt{k_B T}$. Note that the structure of Eq.(17) is identical to our previous estimate of tertiary contact formation time used to interpret kinetics of loop formation in proteins $\tau(n) \approx \frac{\langle R_n^2 \rangle}{D_0 P(n)}$ where n is the loop length, $\langle R_n^2 \rangle$ is the mean square distance between the two residues, D_0 is an effective monomer diffusion constant, and $P(n)$ is the loop formation probability.[8,34,57] The differences between the two lie in the numerical prefactor inside $\{\ldots\}$. In addition, in the dimensional argument used to obtain $\tau(n)$ we used $\langle R_n^2 \rangle$ instead of l_p^2 that arises in the present theory.

Fig.3 shows that the estimates of looping time using Eq.(17) and the results of simulations for the same set of parameters are in excellent agreement when $\xi \approx 7.3$ (see the next section for details of the simulations). First, τ_c increases and converges to the finite value with the increasing size of end tails (decreasing s/l) and this trend manifests itself as the chain gets

stiffer and *shorter* (small s) (see Refs.[30] and[29]). The inset shows that, at $s = 3$, τ_c increases by a factor of ~ 1.5 when the total contour length of the dangling end is 5 times longer than the contour length of the loop. Second, τ_c is a minimum (τ_c^{min}) when the contour length between loop formation sites, $|s_1 - s_2|$, is around $(3 \sim 4)\, l_p^{(0)}$ and τ_c^{min} shifts towards the large s value with the increasing size of dangling ends. Note that when the loop size becomes large ($s > 6$) τ_c does not depend on the length of the dangling ends. In non-interacting Gaussian chains the equilibrium distribution of any two segment along the chain is always Gaussian. In this case, the presence of the dangling ends does not affect the chain statistics.

V.   SIMULATION OF LOOP CLOSURE DYNAMICS

To check the validity of the theoretical estimates for loop closure time we performed simulations using a coarse-grained model for ds-DNA. The simulation procedure is identical to the one used by Podtelezhnikov and Vologodskii (see details in Ref.[58]). Because the time scale of $\sim ms$ is computationally difficult to accomplish, even using Brownian dynamics (BD), we use a coarse-grained model of ds-DNA by choosing the pitch of the helix (10 base pairs with the diameter $l_0 = 3.18nm$) as a building block of a ds-DNA chain. The energy for a worm-like chain, that is appropriate for ds-DNA, is taken to be the sum of the bending rigidity (E_b) term and the chain connectivity (E_s) term, which respectively are given by

$$E_b = \alpha RT \sum_{i=1}^{N-1} \theta_i^2 \tag{18}$$

and

$$E_s = \frac{\beta RT}{l_0^2} \sum_{i=1}^{N} (l_i - l_0)^2. \tag{19}$$

where T is the temperature, R is the gas constant, θ_i is the i-th bond angle, l_i is the i-th bond length. By choosing the parameters $\alpha = 7.775$ and $\beta = 50$ one can get the typical persistence length of $50nm$ for ds-DNA.

Despite the simplification in the energy function computation of the looping time through direct BD simulation still is prohibitively difficult. From Eq.(17) it is clear that the loop formation time can be even up to $\mathcal{O}(1)\ sec$ for certain values of l_p and L. To overcome this problem,

Podtelezhnikov and Vologodskii used the relation between the equilibrium probability of loop formation and the loop closure and opening times,

$$P(r_0; |s_2 - s_1|, L) = \frac{\tau_o}{\tau_{IC} + \tau_o}. \tag{20}$$

τ_o is the loop opening time of the closed loop. In general $\tau_o \ll \tau_{IC}$. This observation enables us to perform direct BD simulation for the loop dissociation rather than loop closure. Since $P(r_0; |s_2 - s_1|, L)$ is normally very small for small r_0 ($r_0 = 5nm$), there is a sampling problem. However, $P(r_0; |s_2 - s_1|, L)$ can be found using the Markov relation

$$P(r_0; |s_2 - s_1|, L) = \prod_{i=1}^{n-1} P(r_i | r_{i+1}) \tag{21}$$

where $P(r_i | r_{i+1})$ is the conditional probability that conformations with $r < r_i$ in the subset of conformations with $r < r_{i+1}$. To obtain $P(r_i | r_{i+1})$ we performed $(n-1)$-Monte Carlo samplings using the pivot algorithm[59] by iteratively adjusting the interval of the end-to-end (or interior-to-interior) distance $r_0 < r_1 < \cdots < r_n$ such that $P(r_i | r_{i+1}) \sim 0.2$.

The results of our simulations for $P(r_0; L)$ and τ_o for both end-to-end contact and contact between segments are shown in Fig.4. Note that there is minor difference between the contact probabilities of the end-to-end and of the interior-to-interior ($|s_2 - s_1|/L = 0.5$) segments whereas the loop opening dynamics for the chain with dangling ends is slower than the case without dangling ends by about $\sim 50\%$. The independence of $P(r_0; |s_2 - s_1|, L)$ justifies the approximation, $G_0(s) \rightarrow G(0, s, l)$, used in obtaining Eq.(17). The values of τ_c and τ_{IC} can be computed knowing $P(r_0; |s_2 - s_1|, L)$ and τ_o. The results, which are shown in Fig.3, are in excellent agreement with theory.

VI.   INTERIOR LOOPING DYNAMICS IN SEMIFLEXIBLE CHAIN WITH VARIABLE PERSISTENCE LENGTH

In many cases stiffness of the loop, which is involved in interior looping, is different from the overall persistence length of the chain. A simple example is the formation of a β-hairpin in peptides. In this case, the stiffness of the loop is typically less than the β-strands. If the

β-hairpin-forming polypeptide chain is treated as a WLC then it is characterized by three fragments, namely, the loop with a persistence length l_{p2} and the strands whose persistence lengths are l_{p1} and l_{p3}. Such a variable persistence length WLC is also realized in the DNA-RNA-DNA construct in which $l_{p2} \approx 10$ Å and $l_{p1} \approx l_{p3} \approx 500$ Å. Constructs consisting of three WLC fragments are also used routinely in laser optical tweezer experiments.

The Kramers type theory, used to calculate interior looping in WLC with uniform l_p, can be adopted to compute τ_{IC} in WLC with variable l_p. The mean field equilibrium distribution function $P(r)$ (with $r = R_{12}/|s_2 - s_1|$), which is needed to calculate τ_{IC} is (see Appendix III for the derivation)

$$P(r) = \frac{4\pi C r^2}{1 - \frac{L_2}{L}r^2} \exp\left(-\frac{3t}{4}\frac{1}{1 - \frac{L_2}{L}r^2}\right) \tag{22}$$

where $L_2 = |s_2 - s_1|$ is the contour length of the loop part of WLC with persistence length l_{p2}, L is the total contour length of the chain, $t = L/l_0^{eff}$. The effective persistence length of the WLC, consisting of three segments $0 < s < s_1$ with l_{p1}, $s_1 < s < s_2$ with l_{p2}, and $s_2 < s < L$ with l_{p3} is

$$l_p^{eff} = \left(\sum_{i=1}^{3} \frac{L_i}{L}\frac{1}{\sqrt{l_{pi}}}\right)^{-2}. \tag{23}$$

In the mean field approximation, the WLC in which l_p varies along the contour in a discrete manner, is equivalent to a WLC with an effective persistence length. It follow from Eq.23 that the effective persistence length is determined by the smallest l_{p_i}.

Consider the simplest case $L_2 = L_{int}$, $L_1 = L_3 = \frac{L - L_{int}}{2}$, $l_{p1} = l_{p3} = l_p^H$, and $l_{p2} = l_p^{(0)}$. In this case,

$$\frac{l_p^{eff}}{l_p^{(0)}} = \frac{1}{\left[(1-x)\sqrt{\frac{l_p^H}{l_p^{(0)}}} + x\right]} \tag{24}$$

where $x = L_{int}/L$. It follows from Eq.24 that if $l_p^H > l_p^{(0)}$ (handle is stiffer than the loop) then $l_p^{eff} > l_p^{(0)}$. Because in the Kramers description τ_{IC} is controlled by l_p^{eff} we expect that interior looping time is greater than τ_{IC} for a chain with uniform l_p. In the opposite limit, $l_p^H < l_p^{(0)}$, (handle is softer than the loop) $l_p^{eff} < l_p^{(0)}$. Consequently, attaching a soft handles should enhance the rate of interior looping.

The interior looping kinetics for a WLC copolymer for different values of the loop and handle persistence lengths are shown in Fig.5. In accord with the arguments given above, we find

that when $l_p^H/l_p^{(0)} = 2$ (stiff handles) τ_{IC} increases substantially compared to $\tau_{IC}(\equiv \tau_{IC}^o)$ for $l_p^H/l_p^{(0)} = 1$ for all values of $|s_2 - s_1|/l_p^{(0)}$ (Fig.5). Similarly, when the handle is softer than that of the loop, τ_{IC} decreases appreciably compared to τ_{IC}^o which is the interior looping the case when the chain has uniform stiffness. In the interesting regime of $|s_2 - s_1|/l_p^{(0)} \approx (2-4)$ we predict a dramatic increase in τ_{IC} compared to τ_{IC}^o when $l_p^H/l_p^{(0)} > 1$ and a substantial decrease in τ_{IC}/τ_{IC}^o when $l_p^H/l_p^{(0)} < 1$ (see inset in Fig.5). Thus, stiff handles retard interior loop kinetics whereas soft handle enhances rates of interior loop formation.

VII.   EFFECT OF MONOMER-MONOMER INTERACTION ON LOOP CLOSURE TIMES

Majority of the recent experiments on dynamics of loop formation have been analyzed using simple polymer models that do not explicitly consider interaction between monomer segments. In a number of cases there are physical interactions between monomers. For instance, DNA is charged and the interaction between the monomers can be approximately described using the short ranged Debye-Hückel potential. Similarly, solvent-mediated interactions also arise especially when considering proteins. For these reasons it becomes necessary to consider an interplay between chain stiffness, entropic fluctuations of the polymer and nonlocal interaction between monomer segments.

The non-linear problem, that arises from the constraint $\mathbf{u}^2(s) = 1$, in a non-interacting semiflexible chain is further exacerbated when interactions between monomers are taken into account.[60] To circumvent this problem we assume that the effect of intra-chain interaction is to only alter the effective persistence length. We compute the loop closure kinetics using Eq.(17) with a renormalized persistence length that explicitly depends on the nature of the interaction between the monomers. This approximation is in the same spirit as the local equilibrium assumption used in this study.

To calculate the renormalized persistence length $l_p^{(R)}$ we follow the procedure due to Hansen and Podgornik[61] who used a mean-field weight (similar to $\Psi_{MF}(\mathbf{u}(s))$ in the presence of non-local interaction, $V(\mathbf{r}(s) - \mathbf{r}(s'))$ between monomers s and s'. The standard field-theoretic procedure is to use the Hubbard-Stratonovich transformation via auxiliary fields to eliminate the non-

Markovian nature of $V(\mathbf{r}(s) - \mathbf{r}(s'))$. Using stationary phase approximation to evaluate the optimal values of the auxiliary fields they[61] obtained an expression for $l_p^{(R)}$ for arbitrary potential $V(\mathbf{r}(s) - \mathbf{r}(s'))$. In our applications, we assume that the charged monomers interact via the screened Coulomb interaction $V(r) = \frac{k_B T l_B}{A^2} \frac{e^{-\kappa r}}{r}$ where l_B is the Bjerrum length ($e^2/4\pi\epsilon_0 k_B T = l_B$), and A is the effective separation between charges on the monomer. The use of Debye-Hückel potential is appropriate when considering ds-DNA in monovalent (Na$^+$) counterions. With this choice of $V(r)$ the renormalized persistence length becomes $l_p^{(R)} = l_p^{(0)} + \delta l_p^{(R)}$, and is given by[61]

$$l_p^{(R)} = l_p^{(0)} + \overline{V}_1 (l_p^{(R)})^2 I(L/l_p^{(R)}, \xi) \tag{25}$$

with

$$I(L/l_p^{(R)}, \overline{\xi}) = \int_0^{L/l_p^{(R)}} dz z^4 e^{-\sqrt{\overline{B}}/\overline{\xi}} \left(\frac{1}{\overline{\xi}\overline{B}} + \frac{1}{\overline{B}^{3/2}} \right) \tag{26}$$

where $\overline{V}_1 = l_B/(12\sqrt{2}dA^2)$, $\overline{B}(z) = z - 1 + e^{-z}$, and $\overline{\xi} = 1/(\sqrt{2}\kappa l_p^{(R)})$. The integral $I(L/l_p^{(R)}, \overline{\xi})$ has different asymptotic behavior depending on the two parameters $\kappa l_p^{(R)}$ and $L/l_p^{(R)}$. (i) If $l_p^{(R)} > \kappa^{-1}$ ($\overline{\xi} < 1$), i.e., the persistence length is greater than the screening length, then the contribution due to electrostatic interaction can be treated perturbatively. In this case the upper limit of the integral in Eq.(26) is effectively set to infinity. We find $\overline{B}(z) \approx z^2$ and $I(L/l_p^{(R)}, \overline{\xi}) \sim \overline{\xi}^2$, which is also small since $\overline{\xi} < 1$. Therefore, $\delta l_p^{(R)} \sim \frac{l_B}{\kappa^2 A^2}$, which coincides with the OSF result.[62,63] For electrostatic contribution to persistence length of a polyelectrolyte chain the limit $l_p^{(R)} > \kappa^{-1}$ is most appropriate for DNA . (ii) if $l_p^{(R)} < \kappa^{-1}$ ($\overline{\xi} > 1$), i.e., the persistence length is smaller than the screening length, there is substantial interaction between the chain segments beyond the length scale of $l_p^{(R)}$. We believe this situation is difficult to be realized in experiments involving biopolymers. In this case, the integral up to $z = \overline{\xi}^2$ becomes important, $\overline{B}(z) \sim z$ and $I(L/l_p^{(R)}, \overline{\xi}) \sim \overline{\xi}^7$. Therefore, $\delta l_p^{(R)} \sim \frac{l_B}{\kappa^7 A^2}$.

To calculate loop closure times the renormalized persistence length $l_p^{(R)}$ is numerically computed for each parameter set (contour length L, inverse screening length κ^{-1}) and we use τ_c (Eq.(17)) with $l_p^{(0)} \rightarrow l_p^{(R)}$. For the ds-DNA in the monovalent salt solution (concentration c) the parameters in the semiflexible chain model are $l_B = 7.1$Å, $A = 1.7$Å and $\kappa = \sqrt{8\pi l_B c}$. The results for τ_c and τ_{IC} are plotted in Figs.6, 7. First, the cyclization times are computed as a function of L at various salt concentrations (Fig.6). We find that τ_c shows a dramatic increase as

c is varied at small values of $L/l_p^{(0)}$. The electrostatic repulsion retards loop closure times, as the salt concentrations (strong inter-segment repulsion) decrease (Fig.6). Because of the interplay between bending rigidity and chain entropy τ_c has a minimum at $t = t^*$. The value of t^* shifts from $t^* = 3$ to $t^* = 6$ as c decreases. The inset in Fig.6 shows that there is practically no change in τ_c at $t = 3$ if $c \gtrsim 100 \; mM$, which is near the physiological concentration ($150 \; mM \; \mathrm{Na^+}$). In this range of c the electrostatic contribution to the persistence length is small so that $l_p^{(R)}$ is almost the same as $l_p^{(0)}$. Note that the c corresponding to the condition $l_p \kappa = 1$ for $l_p = 50 \; nm$ is $c \approx 40 \; mM$.

The dependence of τ_{IC}, which examines the effect of the dangling ends, at high and low concentrations and various values of $|s_2 - s_1|/L$ on $|s_2 - s_1|$ is shown in Fig.7. The insets of Fig.7 show an increased time scale at low salt concentration ($10 \; mM$) compared with high salt concentration ($500 \; mM$). When τ_{IC} for $|s_2 - s_1|/L = 0.2$ is compared with $|s_2 - s_1|/L = 1$, τ_c increases by a factor of ~ 2.7 at $c = 10 \; mM$ whereas the increase is about a factor of ~ 1.5 for $c = 500 \; mM$. The effect of dangling ends on loop formation dynamics manifests itself more clearly at low salt concentration when electrostatic repulsion is prominent and at small ratio of $|s_2 - s_1|/l_p^{(0)}$.

VIII.   DEPENDENCE OF $l_p^{(R)}$ ON THE MONOMER-MONOMER INTERACTION POTENTIAL

In our theory τ_c and τ_{IC} can be determined provided $l_p^{(R)}$ and the distance distribution functions are known. To examine the variation of $l_p^{(R)}$ on the nature of monomer-monomer interactions we have computed $l_p^{(R)}$ for different potentials $V(r)$. Sets of equilibrium conformations of 50-mer bead-spring model (see Eq.(18),(19)) are generated with different bending rigidity, κ_b, and with different non-local potentials $V(r) = 1/r^\alpha (\alpha = 1, 2, 4, 6, 12)$. In each case the effective persistence length $l_p^{(R)}$ is computed by $\frac{1}{1-\cos \langle \theta \rangle}^{64}$ where $\langle \theta \rangle$ is the ensemble average of the angle formed by three consecutive beads. We show the simulated radial distribution function and the effective persistence lengths for different κ_b values for the various ranges of nonlocal interaction in Fig.8-(a),(b). The results from radial distribution function (Fig.8-(a)) and the persistence length show that $l_p^{(R)} \approx l_p^{(0)}$ when the interaction is short-ranged, i.e., $\alpha > 4$. When $\alpha < 4$

then the effective interaction between monomers leads to an increase in the persistence length (Fig.8-(b)). These results are consistent with the field theoretical approach by Hansen and Podgornik.[61] Considering that the excluded volume potential is of short range nature (modeled using $\sim r^{-12}$ or delta function) we conclude that the excluded volume effect on the looping dynamics of rigid polymer chain is negligible. Note that the screened electrostatic potential $V(r) = \frac{e^{-\kappa r}}{r}$ can be either a short or a long range potential depending on the value of κ.

These calculations, especially changes in l_p as the range of interaction is altered, explain the reason that a simple WLC model works remarkably well in a number of applications. For example, the response of DNA, RNA,[65,66] and proteins[67] to mechanical force has been routinely analyzed using WLC. Surprisingly, recent analysis of small angle x-ray scattering measurements[45] on ribozymes have shown that the distance distribution function can be quantitatively fit using $P(R_e)$. In these biopolymers the interactions that determine the conformations are vastly different. However, the results in Fig.8-(b) show that as long as these effective interactions are short-ranged l_p should not differ from the bare persistence length. This key result rationalizes the use of WLC in seemingly diverse set of problems.

IX. CONCLUSIONS

In this paper we have used theory and explicit simulations of worm-like chains to examine loop formation dynamics with emphasis on kinetics of contact formation between monomers that are in the interior of the chain. The Kramers theory, adopted to describe looping time scales using the analytically computed potential of force between the contacting (or reacting) groups, gives results that are in quantitative agreement with simulations. The theory[37] for τ_{IC} and τ_c contains one parameter that was introduced to account for the uncertainty in the estimate of the frequency at the transition state (Fig.2) in the intramolecular reaction (Fig.1). The present study also provides a justification for the use of Kramers-like theory in describing looping dynamics by explicit comparison with simulations of semiflexible chains. Although several questions of fundamental theoretical importance remain previous studies, beginning with the pioneering work by WF, and the present study have given a practical analytic formula to analyze most of the

recent experimental data on proteins and DNA. We conclude the paper with a few additional comments.

1. The present work and several previous studies,[15,37,57,58] which have examined the effect of stiffness on looping dynamics, have shown that the rates of cyclization and interior looping must slow down as the loop length becomes small. In other words, there must be a turnover in the plot of k_α ($\alpha = c$ or IC) as s decreases (see Fig.3 in which τ_{IC} as a function of s is shown). For the parameters used in Fig.3 the turnover occurs around $s \approx (3-4)$. Such a crossover has been observed in the cyclization kinetics of DNA and in simulations of worm like chain models. When s is small the time scales for loop formation can be substantially large ($\sim \mathcal{O}(1)\ sec$).

   The effect of stiffness on cyclization rate in disordered peptides has also been emphasized.[15] For the construct $Cys - (Ala - Gly - Gln)_j - Trp$ with j from $1-6$ the stiffness effects are evident at $j \approx 3$.[15] However, these authors did not observe the theoretically predicted turnover in this construct for which the persistence length is estimated to be $l_p \approx 0.7nm$. Using the results in Fig. (3) we predict that the turnover must occur only when the number of peptide bonds is less than about 3. This limit has not been reached in the experiments by Lapidus $et.\ al.$[15] For the construct $(Gly - Ser)_j$ Hudgins $et.\ al.$[27] have clearly observed a turnover when the number of peptide decreases below about 4. The observation of Hudgins $et.\ al.$ is consistent with our prediction that turnover in cyclization rates in disordered polypeptides occurs when $s \lesssim 3$. When the number of residues in the polypeptides chain becomes too small then measuring τ_c using bulky donor-acceptor pairs in FRET experiments is difficult. In this situation other methods[22] could be used.

2. For the parameters used in Fig. (3) the difference between τ_c and τ_{IC} is no more than about a factor of four. However, if charged interactions between monomers become relevant then τ_{IC} can be very different from τ_c. At both low and high values of the salt concentration the τ_c and τ_{IC} can differ by nearly an order of magnitude (see Fig.(6)). These variations are significant because τ_c in polypeptide chains studied thus far varies by less than a factor of ten as the number of residues is varied from 5 to 20. It would be interesting to probe looping dynamics by varying the net charge on polypeptides. We should also stress that as

the salt concentration increases the electrostatic interactions in a high dielectric medium are effectively short ranged. In this case τ_{IC} is determined essentially by the bending rigidity of the backbone (Fig.8).

3. The Kramers based theory for τ_c and τ_{IC} is a convenient way to measure persistence length of polypeptide chains as a function of temperature and denaturant concentration. Recent measurements suggest that l_p depends on urea concentration.[23] More importantly, there appears to be strong sequence effects in τ_c which, at the level of polymer-based theories, must reflect changes in l_p. For example, τ_c for polyproline deviates substantially from ideal chain behavior.[43] Similar measurements of τ_c for other polypeptides along with the simple theory can be used to extract how l_p varies with sequence.

4. The dependence of interior looping time on the ratio of the persistence lengths of loop and the handle shows that in the interesting range of $s \sim (2-4)$ τ_{IC} can be substantially larger than τ_{IC}^o for a chain in which $l_p^H/l_p^{(0)} = 1$ (see inset in Fig.5). This case is directly applicable to β-hairpin formation that is controlled by formation of a loop with persistence length that is less than that of the strands.[20] From Fig.5 it follows that as the stiffness of the loop increases the interior looping time also increases. This conclusion is in accord with explicit simulations of coarse-grained models of β-hairpin formation that showed that enhancement of loop stiffness retards rate of β-hairpin formation.[20]

APPENDIX A

In this Appendix we outline the steps leading to Eq.(6). The distribution function $G(R_{12}, s_1 - s_2)$ is

$$
\begin{aligned}
G(R_{12}, s_1 - s_2) = e^{\lambda L + 2\delta} \int \frac{d^3\mathbf{k}}{(2\pi)^3} \int & d\mathbf{u}_0 d\mathbf{u}_{s_1} d\mathbf{u}_{s_2} d\mathbf{u}_L \\
& \times e^{-\delta \mathbf{u}_0^2} Z(\mathbf{u}_0, \mathbf{u}_{s_1}; s_1) \\
& \times e^{-\frac{\mathbf{k}^2}{4\lambda}|s_2 - s_1| + i\mathbf{k}\cdot\mathbf{R}_{12}} Z\left(\mathbf{u}_{s_1} + \frac{i\mathbf{k}}{2\lambda}, \mathbf{u}_{s_2} + \frac{i\mathbf{k}}{2\lambda}; s_2 - s_1\right) \\
& \times e^{-\delta \mathbf{u}_L^2} Z(\mathbf{u}_{s_2}, \mathbf{u}_L; L - s_2)
\end{aligned}
\tag{A.1}
$$

Using the expression Eq.(4) for $Z(\mathbf{u}_s, \mathbf{u}_0, s)$ and carrying out the integrals over the $\mathbf{u}$ variables Eq.(A.1) becomes

$$G(R_{12}, s_1 - s_2) = e^{\lambda L + 2\delta} \left(\frac{\pi \sinh \Omega s_1}{\Omega_p} \right)^{-3/2} \left(\frac{\pi \sinh \Omega(s_2 - s_1)}{\Omega_p} \right)^{-3/2} \left(\frac{\pi \sinh \Omega(L - s_2)}{\Omega_p} \right)^{-3/2}$$

$$\times \int \frac{d^3\mathbf{k}}{(2\pi)^3} \exp \left(i\mathbf{k} \cdot \mathbf{R}_{12} - \frac{\mathbf{k}^2}{4\lambda} |s_2 - s_1| + \frac{\mathbf{k}^2}{2\lambda^2} \Omega_p (\coth \Omega(s_2 - s_1) - \frac{1}{\sinh \Omega(s_2 - s_1)}) \right)$$

$$\times \left(\frac{\pi}{\det \mathcal{A}} \right)^{3/2} \exp \left(\frac{1}{4} \mathbf{b}^T(\mathbf{k}) \mathcal{A}^{-1} \mathbf{b}(\mathbf{k}) \right) \tag{A.2}$$

where

$$\mathcal{A} = \begin{pmatrix} \delta + \Omega_p \coth \Omega s_1 & -\frac{\Omega_p}{\sinh \Omega s_1} & 0 & 0 \\ -\frac{\Omega_p}{\sinh \Omega s_1} & \Omega_p(\coth \Omega s_1 + \coth \Omega(s_2 - s_1)) & -\frac{\Omega_p}{\sinh \Omega(s_2 - s_1)} & 0 \\ 0 & -\frac{\Omega_p}{\sinh \Omega(s_2 - s_1)} & \Omega_p(\coth \Omega(s_2 - s_1) + \coth \Omega(L - s_2)) & -\frac{\Omega_p}{\sinh \Omega(L - s_2)} \\ 0 & 0 & -\frac{\Omega_p}{\sinh \Omega(L - s_2)} & \delta + \Omega_p \coth \Omega(L - s_2) \end{pmatrix}, \tag{A.3}$$

and

$$\mathbf{b}^T = \frac{i\mathbf{k}}{\lambda} \Omega_p (\coth \Omega(s_2 - s_1) - \frac{1}{\sinh \Omega(s_2 - s_1)}) \begin{pmatrix} 0 & 1 & 1 & 0 \end{pmatrix}. \tag{A.4}$$

The integration with respect to $\mathbf{k}$ leads to

$$G(R_{12}, s_1, s_2) = e^{\lambda L + 2\delta} \left(\frac{\pi \sinh \Omega s_1}{\Omega_p} \right)^{-3/2} \left(\frac{\pi \sinh \Omega(s_2 - s_1)}{\Omega_p} \right)^{-3/2} \left(\frac{\pi \sinh \Omega(L - s_2)}{\Omega_p} \right)^{-3/2}$$

$$\times \left(\frac{\pi}{\det \mathcal{A}} \right)^{3/2} \frac{1}{(2\pi)^3} \left(\frac{4\pi\lambda^2}{Q(s_1, s_2; \lambda, \delta)} \right)^{3/2} e^{-\frac{\lambda^2 R_{12}^2}{Q(s_1, s_2; \lambda, \delta)}} \tag{A.5}$$

where

$$\det \mathcal{A} = \frac{\Omega_p^2 \sinh \Omega L}{\sinh \Omega s_1 \sinh \Omega(s_2 - s_1) \sinh \Omega(L - s_2)} \{ (\delta^2 + \Omega_p^2) + 2\delta\Omega_p \coth \Omega L \}, \tag{A.6}$$

$$Q(s_1, s_2; \lambda, \delta) = |s_2 - s_1| \lambda - 2\Omega_p (\coth \Omega(s_2 - s_1) - \frac{1}{\sinh \Omega(s_2 - s_1)}) + M(s_1, s_2), \tag{A.7}$$

and

$$M(s_1, s_2; \lambda, \delta) = \frac{\Omega_p \sinh \frac{\Omega(s_2 - s_1)}{2} \tanh^2 \frac{\Omega(s_2 - s_1)}{2}}{2\delta\Omega_p \cosh \Omega L + (\delta^2 + \Omega_p^2) \sinh \Omega L}$$

$$\times \{ 4\delta\Omega_p \cosh \frac{\Omega(2L - (s_2 - s_1))}{2} + 2(\delta^2 + \Omega_p^2) \sinh \frac{\Omega(2L - (s_2 - s_1))}{2}$$

$$- (\delta^2 - \Omega_p^2)(\sinh \frac{\Omega(2L - 3s_1 - s_2)}{2} - \sinh \frac{\Omega(2L - s_1 - 3s_2)}{2}) \} \tag{A.8}$$

In limit $L \gg s_2 \gg s_1 \gg 1$, $M(s_1, s_2) \to \Omega_p$ and $Q(s_1, s_2; \lambda, \delta) \to |s_2 - s_1|\lambda - \Omega_p$. As a result of translational symmetry along the chain $G(R, s_1, s_2) = G(R, |s_2 - s_1|)$. Using $\mathcal{F}(\lambda, \delta) \approx -\ln G(R_{12}, s_1, s_2)$ leads to Eq.(6).

APPENDIX B

In obtaining the stationarity condition to evaluate λ and δ we first took the thermodynamic limit ($L \to \infty$) and then calculated the optimal values of λ and δ. It is technically necessary to solve the stationarity condition $\frac{\mathcal{F}}{\partial \lambda} = \frac{\partial \mathcal{F}}{\partial \delta} = 0$ before taking the $L \to \infty$ limit. In this appendix we examine the consequence of taking the thermodynamic limit after solving for optimal values of λ and δ. For simplicity, set $|s_2 - s_1| = L$. The variational equations for λ and δ become

$$\delta + \Omega_p = \frac{3}{2} \tag{B.1}$$

and

$$L\left[\left(\frac{3}{4\Omega_p} - 1 + \frac{R^2}{L^2}\right) - \frac{l_p}{L}\left(\frac{9}{8\Omega_p^2} - \frac{1}{2\Omega_p}\right)\right] = 0 \tag{B.2}$$

From the second relation we find two roots for Ω_p, namely,

$$\Omega_p^{\pm} = \sqrt{\frac{\lambda^{\pm} l_p}{2}} = \frac{3(1 + \frac{1}{t})}{4(1 - r^2)}\left[\frac{1}{2} \pm \frac{1}{2}\sqrt{1 - \frac{12(1 - r^2)}{t(1 + \frac{1}{t})^2}}\right] \tag{B.3}$$

where $t = L/l_0$, $l_0 = \frac{2}{3}l_p$ and $r = R/L$. There are no restrictions on the values of L and l_p in Eq.(7) in which the thermodynamic limit is taken first. However, when the order of operation is exchanged there is a possibility that the two roots are $\Omega_p^{\pm}$ that can be imaginary. For Ω_p^+ we retrieve the same stationary phase condition as Eq.(8) only if $L \gg l_p(t \gg 1)$ The second root $\Omega_p^- = 0$ but this can be discarded since $\lambda \neq 0$. Although there are multiple saddle points, we can always deform the contour such that the contour passes the saddle point with $\mathcal{F}'(\Omega_p^+) = 0$ which satisfies the stationarity condition. In addition, $\frac{1}{12}(10 - t - \frac{1}{t}) < r^2 < 1$ should be always satisfied for Ω_p to be real (Eq.(B.3)).

APPENDIX C

We calculate $P(\mathbf{R}_{12}, |s_2 - s_1|)$ for a semiflexible chain in which the persistence length of the loop is different from that of the dangling ends. The chain can be thought of as a triblock WLC copolymer which is an appropriate model for RNA hairpins or β-hairpins in disordered polypeptides. The interior $\mathbf{R}_{12}$ distribution function is

$$G(\mathbf{R}_{12}; L_1, L_2, L_3) = \langle \delta(\mathbf{R}_{12} - \int_{L_2} \mathbf{u}(s)ds) \rangle_{MF}$$

$$= \frac{\int D[\mathbf{u}(s)]\delta(\mathbf{R}_{12} - \int_{L_2} \mathbf{u}(s)ds)\Psi_{MF}[\mathbf{u}(s)]}{\int D[\mathbf{u}(s)]\Psi_{MF}[\mathbf{u}(s)]}, \tag{C.1}$$

where $\mathbf{u}(s)$ is a tangent vector at position s. Suppose the chain consists of three different parts characterized by the persistence lengths (l_{pi}) and the contour lengths (L_i) with $i = 1, 2, 3$ ($L = \sum_{i=1}^{3} L_i$). The exact weight $\Psi[\mathbf{u}(s)] \propto \exp\left[-\sum_{i=1}^{3} \frac{l_{pi}}{2} \int_{L_i} ds \left(\frac{\partial \mathbf{u}}{\partial s}\right)^2\right] \prod \delta(\mathbf{u}^2(s) - 1)$ is replaced by the mean field weight $\Psi_{MF}[\mathbf{u}(s)]$.

$$\Psi_{MF}[\mathbf{u}(s)] \propto \exp\left[-\sum_{i=1}^{3} \frac{l_{pi}}{2} \int_{L_i} \left(\frac{\partial \mathbf{u}(s)}{\partial s}\right)^2 ds - \sum_{I=1}^{3} \lambda \int_{L_i} (\mathbf{u}^2(s) - 1)ds - \delta[(\mathbf{u}_0^2 - 1) + (\mathbf{u}_L^2 - 1)]\right]. \tag{C.2}$$

As in Appendix A, the Lagrange multipliers λ and δ are used to enforce the local constraint $\mathbf{u}^2(s) = 1$. Following, exactly the procedure outlined in Appendix A we find that in the $L \to \infty$ limit, the analogue of Eq.7 becomes,

$$\mathcal{F}[\lambda, \delta] \approx -(L\lambda + 2\delta)$$

$$+ \sum_{i=1}^{3} \frac{3}{2} \ln \frac{e^{\Omega_i L_i}}{\Omega_{pi}} + \frac{3}{2} \ln\left[(\delta + \Omega_{p1})(\Omega_{p1} + \Omega_{p2})(\Omega_{p2} + \Omega_{p3})(\delta + \Omega_{p3})\right] + \frac{3}{2} \ln \frac{L_2}{\lambda} + \frac{\mathbf{R}_{12}^2 \lambda}{L_2}$$

$$= L\left[\frac{3}{2}\left(\Omega_1 \frac{L_1}{L} + \Omega_2 \frac{L_2}{L} + \Omega_3 \frac{L_3}{L}\right) - \lambda\left(1 - \frac{L_2}{L} \frac{\mathbf{R}_{12}^2}{L_2^2}\right)\right]$$

$$+ \frac{3}{2} \ln\left[\frac{\Omega_{p2}}{\lambda}\left(\frac{\delta}{\Omega_{p1}} + 1\right)\left(\frac{\Omega_{p1}}{\Omega_{p2}} + 1\right)\left(\frac{\Omega_{p3}}{\Omega_{p2}} + 1\right)\left(\frac{\delta}{\Omega_{p3}} + 1\right)\right] + \frac{3}{2} \ln L_2 - 2\delta. \tag{C.3}$$

where $\Omega_i = \sqrt{2\lambda/l_{pi}}$ and $\Omega_{pi} = \Omega_i l_{pi}/2$. The major contribution to the integral comes from the sets of λ and δ which pass the saddle point of a stationary phase contour on $\mathbf{Re}\{\mathcal{F}\}$ plane. The stationary condition for λ by taking derivative with respect to λ by retaining the leading term in L, which leads to

$$\lambda^{1/2} = \frac{3}{4(1 - \frac{L_2}{L}r^2)}\left(\sum_{i=1}^{3} \sqrt{\frac{2}{l_{pi}}} \frac{L_i}{L}\right). \tag{C.4}$$

where $r = \frac{R_{12}}{L_2}$. Similarly, the condition for δ results

$$\frac{1}{\delta + \Omega_{p1}} + \frac{1}{\delta + \Omega_{p3}} = \frac{4}{3}. \tag{C.5}$$

Substituting Eq.C.4 and C.5 to $G(R) \approx \exp\left(-\mathcal{F}[\mathbf{R}; \lambda, \delta]\right)$ gives the desired distribution function,

$$G(R_{12}; \{l_{pi}\}, \{L_i\}) = \frac{N}{(1 - \frac{L_2}{L} r^2)^{9/2}} \exp\left[-\frac{9}{8} \frac{L}{1 - \frac{L_2}{L} r^2} \left(\sum_{i=1}^{3} \sqrt{\frac{1}{l_{pi}}} \frac{L_i}{L} \right)^2 \right]. \tag{C.6}$$

The effective persistence length in the mean field approximation is

$$l_p^{eff} = \left(\sum_{i=1}^{3} \frac{L_i}{L} \frac{1}{\sqrt{l_{pi}}} \right)^{-2}. \tag{C.7}$$

The result derived for a triblock WLC can be generalized into N-block WLC with persistence lengths $\{l_{pi}\}$ and contour lengths $\{L_i\}$. The distribution function is

$$P(r) = \frac{4\pi C r^2}{(1 - \frac{L_2}{L} r^2)^{9/2}} \exp\left(-\frac{3t}{4} \frac{1}{1 - \frac{L_2}{L} r^2} \right) \tag{C.8}$$

where $t = L/l_o^{eff}$, $l_o^{eff} = \frac{2}{3} \left(\sum_{i=1}^{N} \frac{L_i}{L} \frac{1}{\sqrt{l_{pi}}} \right)^{-2}$, $r = R_{12}/|s_2 - s_1|$ and C is a normalization constant.

[1] Flory, P. J. (1971) *Principles of Polymer Chemistry* (Cornell University Press, Ithaca).

[2] Winnik, M. A. (1986) in *Cyclic Polymers* (Elsvier, New York), edited by J. A. Semylen.

[3] Wilemski, G. & Fixman, M. (1974) *J. Chem. Phys.* **60**(3), 866–877.

[4] Pastor, R. W., Zwanzig, R., & Szabo, A. (1996) *J. Chem. Phys.* **105**(9), 3878–3882.

[5] Doi, M. (1975) *Chem. Phys.* **9**, 455–466.

[6] Szabo, A., Schulten, K., & Schulten, Z. (1980) *J. Chem. Phys.* **72**(8), 4360–4357.

[7] Battezzati, M. & Perico, A. (1981) *J. Chem. Phys.* **74**, 4527–4532.

[8] Thirumalai, D. (1999) *J. Phys. Chem. B* **103**, 608–610.

[9] Wilemski, G. & Fixman, M. (1973) *J. Chem. Phys.* **58**, 4009–4019.

[10] Cumberti, C. & Perico, A. (1984) *Prog. Polymer. Sci.* **10**, 271.

[11] Rey, A. & Freire, J. J. (1991) *Macromolecules* **24**, 4673.

[12] Podtelezhnikov, A. & Vologodskii, A. (1997) *Macromolecules* **30**, 6668–6673.

[13] Hagen, S. J., Hofrichter, James, & Eaton, W. A. (1997) *J. Phys. Chem. B* **101**, 2352–2365.

[14] Shore, D., Langowski, J., & Baldwin, R. L. (1981) *Proc. Natl. Acad. Sci.* **78**(8), 4833–4837.

[15] Lapidus, L. J., Steinbach, P. J., Eaton, W. A., Szabo, A., & Hofrichter, J. (2002) *J. Phys. Chem. B* **106**, 11628–11640.

[16] Du, Q., Smith, C., Shiffeldrim, N., Vologodskaia, M., & Vologodskii, A. (2005) *Proc. Natl. Acad. Sci.* **102**, 5397–5402.

[17] Tinoco, Jr., I. & Bustamante, C. (1999) *J. Mol. Biol.* **293**, 271–281.

[18] Bonnet, G., Krichevsky, O., & Libchaber, A. (1998) *Proc. Natl. Acad. Sci.* **95**, 8602–8606.

[19] Kuznetsov, S. V., Shen, Y., Benight, A. S., & Ansari, A. (2001) *Biophys. J.* **81**, 2864–2875.

[20] Klimov, D. K. & Thirumalai, D. (2000) *Proc. Natl. Acad. Sci.* **97**, 2544–2549.

[21] Bieri, O., Wirz, J., Hellrung, B., Schutkowski, M., Drewello, M., & Kiefhaber, T. (1999) *Proc. Natl. Acad. Sci.* **96**, 9597–9601.

[22] Chang, I. J., Lee, J. C., Winkler, J. R., & Gray, H. B. (2003) *Proc. Natl. Acad. Sci.* **100**, 3838–3840.

[23] Bodenreider, C. & Kiefhaber, T. (2005) *J. Mol. Biol.* **345**, 153–162.

[24] Merlitz, H., Rippe, K., Klenin, K. V., & Langowski, J. (1998) *Biophys. J.* **74**(2), 773–779.

[25] Halford, S. E., Welsh, A. J., & Szczelkun, M. D. (2004) *Ann. Rev. Biophys. Biomol. Struct.* **33**,

1–24.

[26] Hagen, S. J., Hofrichter, J., Szabo, A., & Eaton, W. A. (1996) *Proc. Natl. Acad. Sci.* **93**, 11615–11617.

[27] Hudgins, R. R., Huang, F., Gramlich, G., & Nau, W. M. (2002) *J. Am. Chem. Soc.* **124**, 556–564.

[28] Cloutier, T. & Widom, J. (2004) *Mol. Cell.* **14**, 355–362.

[29] Lee, S. & Winnik, M. A. (1997) *Macromolecules* **30**, 2633–2641.

[30] Perico, A. & Beggiato, M (1990) *Macromolecules* **23**, 797–803.

[31] Friedman, B. & O'Shaughnessy, B. (1991) *J. Phys. II* **1**, 471–486.

[32] Camacho, C. J. & Thirumalai, D. (1995) *Proteins: Structure, Function and Genetics* **22**, 28–40.

[33] Jones, C. M., Henry, E. R., Hu, Y., Chan, C. K., Luck, S. D., Bhuyan, A., Roder, H., Hofrichter, J., & Eaton, W. A. (1993) *Proc. Natl. Acad. Sci.* **90**, 11860–11864.

[34] Guo, Z. & Thirumalai, D. (1995) *Biopolymers* **36**, 83–102.

[35] Thirumalai, D. & Hyeon, C. (2005) *Biochemistry* **44**(13), 4957–4970.

[36] des Cloizeaux, J. (1974) *Phys. Rev. A* **10**, 1665–1669.

[37] Jun, S., Bechhoefer, J., & Ha, B.-Y. (2003) *Europhys. Lett.* **64**(3), 420–426.

[38] Wilemski, G. & Fixman, M. (1974) *J. Chem. Phys.* **60**(3), 878–890.

[39] de Gennes, P. G. (1982) *J. Chem. Phys.* **76**(6), 3316–3321.

[40] des Cloizeaux, J. (1980) *J. Phys.* **41**, 223–238.

[41] Jacobson, H. & Stockmayer, W. H. (1950) *J. Chem. Phys.* **18**, 1600–1606.

[42] Portman, J. J. (2003) *J. Chem. Phys.* **118**(3), 2381–2391.

[43] Huang, F. & Nau, W. M. (2005) *Res. Chem. Interm.* **31**, 717–726.

[44] Laurence, T. A., Kong, X., Jäger, M., & Weiss, S. (2005) *Proc. Natl. Acad. Sci.* **102**, 17348–17353.

[45] Caliskan, G., Hyeon, C., Perez-Salas, U., Briber, R. M., Woodson, S. A., & Thirumalai, D. (2005) *Phys. Rev. Lett.* **95**, 268303.

[46] Hyeon, C., Dima, R. I., & Thirumalai, D. Size, shape and fleixibility of RNA structures, unpublished.

[47] Dua, A. & Cherayil, B. J. (2002) *J. Chem. Phys.* **116**(1), 399–409.

[48] Wilhelm, J. & Frey, E. (1996) *Phys. Rev. Lett.* **77**(12), 2581–2584.

[49] Bhattacharjee, J.K., Thirumalai, D., & Bryngelson, J.D. (1997) *cond-mat/9709345*.

[50] Thirumalai, D. & Ha, B.-Y. (1998) (Academic Press, San Diego), chapter I. Statistical Mechanics of Semiflexible Chains: A Mean Field Variational Approach in *Theoretical and Mathematical Models*

in Polymer Research, pp. 1–35, edited by A. Grosberg.

[51] Schuler, B., Lipman, E. A., Steinbach, P. J., Kumke, M., & Eaton, W. A. (2005) *Proc. Natl. Acad. Sci.* **102**, 2754–2759.

[52] Murphy, M. C., Rasnik, I., Cheng, W., Lohman, Timothy M., & Ha, T. (2004) *Biophys. J.* **86**, 2530–2537.

[53] Lagowski, J.B., Noolandi, J., & Nickel, B. (1991) *J. Chem. Phys.* **95**, 1266–1269.

[54] Ha, B.-Y. & Thirumalai, D. (1995) *J. Chem. Phys.* **103**(21), 9408–9412.

[55] Shimada, J. & Yamakawa, H. (1984) *Macromolecule* **17**, 689–698.

[56] Ringrose, L., Chabanis, S., Angrand, P., Woodroofe, C., & Stewart, A.F. (1999) *EMBO* **18**(23), 6630–6641.

[57] Camacho, C. J. & Thirumalai, D. (1995) *Proc. Natl. Acad. Sci.* **92**, 1277.

[58] Podtelezhnikov, A. A. & Vologodskii, A. V. (2000) *Macromolecules* **33**, 2767–2771.

[59] Bishop, M., Clarke, J. H. R., Rey, A., & Freire, J. J. (1991) *J. Chem. Phys.* **95**, 4589–4592.

[60] Lee, N. K. & Thirumalai, D. (1999) *Europhys. J. B.* **12**, 599–605.

[61] Hansen, P. L. & Podgornik, R. (2001) *J. Chem. Phys.* **114**(19), 8637–8648.

[62] Odijk, T. (1977) *J. Polym. Sci., Polym. Phys.* **15**, 477–483.

[63] Skolnick, J. & Fixman, M. (1977) *Macromolecules* **10**, 944–948.

[64] Lee, N. & Thirumalai, D. (2004) *Biophys. J.* **86**, 2641–2649.

[65] Liphardt, J., Onoa, B., Smith, S.B., Tinoco, Jr., I., & Bustamante, C. (2001) *Science* **292**, 733–737.

[66] Onoa, B., Dumont, S., Liphardt, J., Smith, S.B., Tinoco, Jr., I., & Bustamante, C. (2003) *Science* **299**, 1892–1895.

[67] Rief, M., Gautel, M., Oesterhelt, F., Fernandez, J. M., & Gaub, H. E. (1997) *Science* **276**, 1109–1111.

FIGURE CAPTIOINS

Figure 1 : Loop formation in semiflexible polymer chains: (a) Cyclization event. (b) Interior contact (IC) formation between monomers s_1 and s_2 in the chain interior. The segment length s_1 and $L - s_2$ are referred to either as dangling ends or handles in the text.

Figure 2 : Comparison between the interior distance distribution functions for different size of the loops ($\frac{|s_2 - s_1|}{L}$ =1, 0.8, 0.5). The value of $|s_2 - s_1|/L = 1$ corresponds to end-to-end distribution function. The flexible chain limit ($t = 10$) is on the left and the right panel is for stiff chains ($t = 2$). The panel on the right shows a sketch of the effective potential $F(r) = -k_B T \log P(r)$ for the case $|s_2 - s_1|/L = 0.8$. The wells at R_b and the barrier top R_t are highlighted.

Figure 3 : Plots of loop formation time (τ_{IC}) obtained using Eq.(17) as a function of the distance between site s_1 and s_2 expressed in terms of $l_p^{(0)}$ for various size of dangling ends expressed by $s/l(= |s_1 - s_2|/L)$. Here $l = L/l_p^{(0)}$ and $s = |s_2 - s_1|/l_p^{(0)}$. The specific values for parameters are $l_p^{(0)} = 50\ nm$, $D = 2D_0 = 1.54 \times 10^{-11} m^2/s$ and $\alpha = 0.1$. The insets are for the loop closing time at $|s_1 - s_2|/l_p^{(0)} = 3$ as a function of s/l. For two sites separated by $|s_1 - s_2|$ along the chain the longer loop closing time is expected if the sites of interest are connected by long dangling ends. If the separation is much larger than the persistence length ($|s_1 - s_2| >> l_p^{(0)}$) the effect of dangling end on loop closing time vanishes. The results of simulations for the same set of parameters are shown in symbols. The excellent agreement between theory and simulations validates the assumptions leading to Eq.(17).

Figure 4 : Plot of looping probability, $P(r_0; S)$, (left) and dissociation time, τ_o, (right) as a function of interior loop contour length S. The capture radius $r_0 = 5nm$. The simulations are performed for both end-to-end ($S/L = 1$) (black circle) and interior-to-interior ($S/L = 0.5$) (red triangle). The parameters of the semiflexible chains are the same as in Fig.3. Note the loop dissociation time is much shorter than τ_{IC}.

Figure 5 : Interior looping time (τ_{IC}) as a function of the reduced distance between sites s_1 and s_2 for the dangling ends of $s/l = 0.3$ for WLC with variable persistence lengths ($l_p^H/l_p^{(0)} = 2$ in green, triangle up, $l_p^H/l_p^{(0)} = 1$ in red thick line, and $l_p^H/l_p^{(0)} = 0.5$ in blue, triangle down). The physical situation corresponds to Fig.1 in which the persistence length of the two handles $(0, s_1)$

and (s_2, L) is l_p^H and the contour lengths of three segments are identical. For comparison the cyclization time (τ_c with $s/l = 1$ and the chain persistence $l_p^{(0)}$) as a function of $s(= l = L/l_p^{(0)})$ is shown in black dashed line. The inset shows the ratios of looping times (τ_{IC}) for the chain with variable persistence length with respect to the looping time (τ_{IC}^o) of the chain with uniform persistence. The up triangle in green is for the ratio between $l_p^H/l_p^{(0)} = 2$, $s/l = 0.3$ and $l_p^H/l_p^{(0)} = 1$, $s/l = 0.3$ (stiff handle), and the dpwm triangle down in blue is for the ratio between $l_p^H/l_p^{(0)} = 0.5$, $s/l = 0.3$ and $l_p^H/l_p^{(0)} = 1$, $s/l = 0.3$ (soft handle).

Figure 6 : Plots of cyclization time (τ_c) as a function of L (expressed in terms of $l_p^{(0)}$) for various salt concentrations. The same parameters ($l_p^{(0)}$, D, α) with those in Fig.3 are used. The inset shows τ_c at $L/l_p^{(0)} = 3$ as a function of c. The cyclization time τ_c increases sharply below $c \lesssim 50mM$.

Figure 7 : Interior looping time (τ_{IC}) as a function of the reduced distance between site s_1 and s_2 for various size of dangling ends under two salt concentrations. The length of the dangling end is given by the parameter $s/l(= |s_1 - s_2|/L)$. The same parameters ($l_p^{(0)}$, D, α) as in Fig.3 are used. The values of the salt concentration (10mM, 500mM) are explicitly shown. At each value of c the different curves correspond to distinct values of s/l. The values of s/l range from 0.2 to 1. The inset shows τ_{IC}, at the two values of c, as a function of s/l for $|s_2 - s_2|/l_p^{(0)} = 3$ (the vertical dashed line).

Figure 8 : (a) End-to-end radial distribution function for a semiflexible chain with bending rigidity ($\kappa_b = 3, 10, 20, 50$, and 100 in unit of k_BT) for various non-local interaction potentials between monomers. The form of the potential is $V(r) = r^{-\alpha}$, with $\alpha = 1, 2, 4, 6, 12$: Results are obtained using Brownian dynamics simulation using the energy function $E_{WLC} = E_b + E_s + \sum_{i<j} V(r_{ij})$. (b) The effective persistence length for various potentials $V(r)$ at different values of the bare bending rigidities.
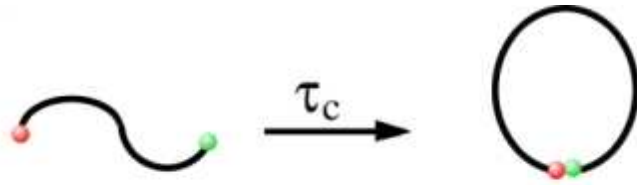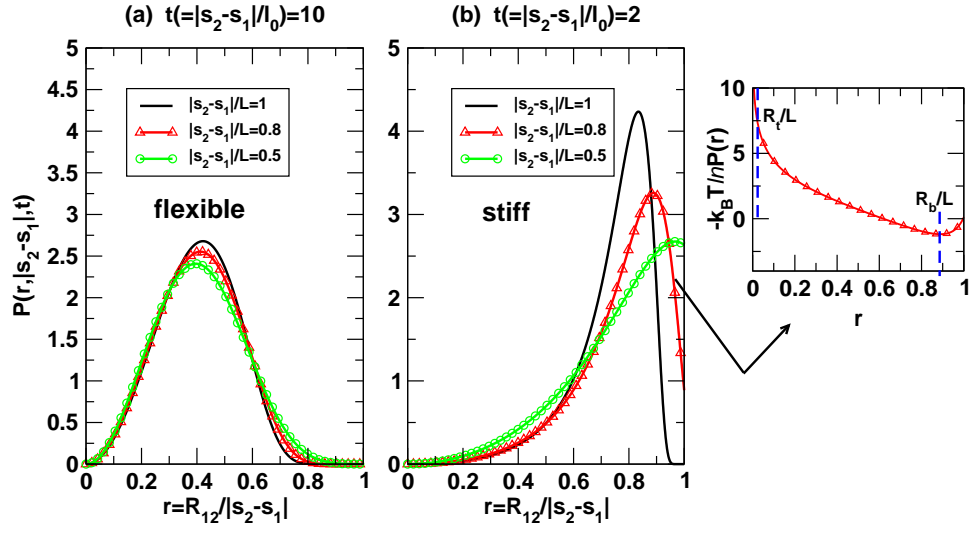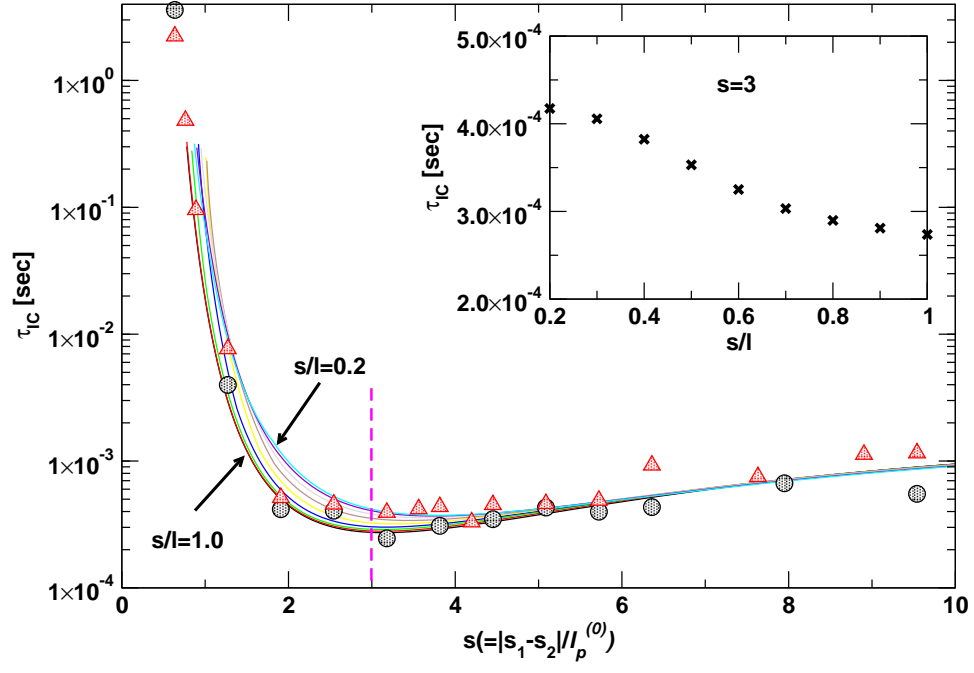
(a)
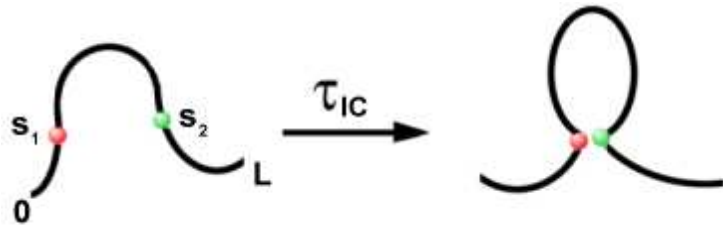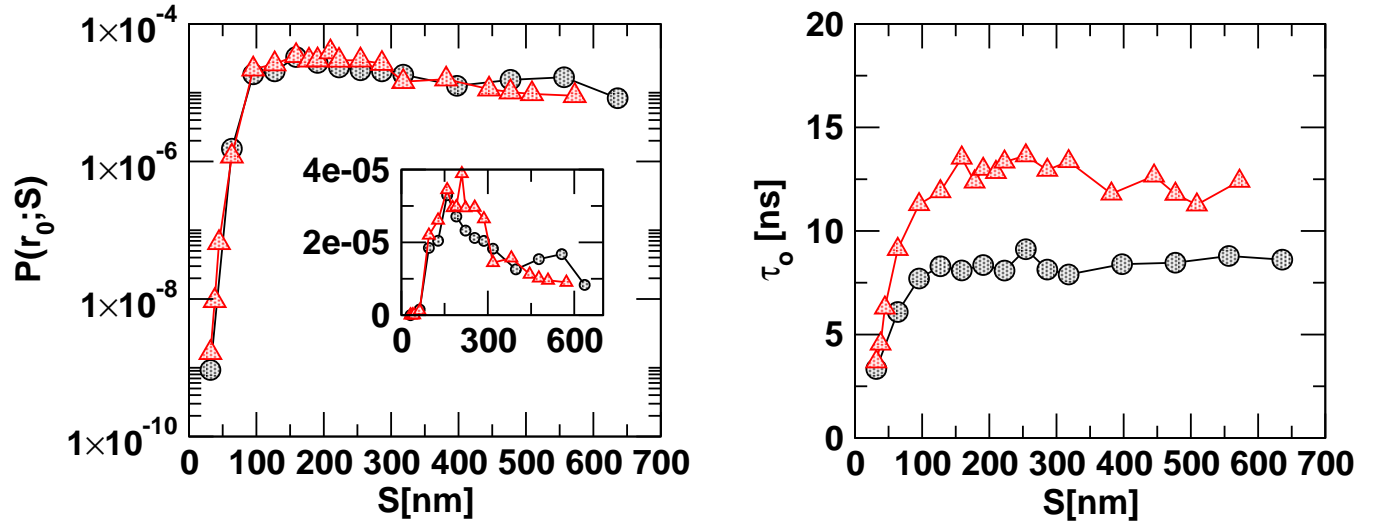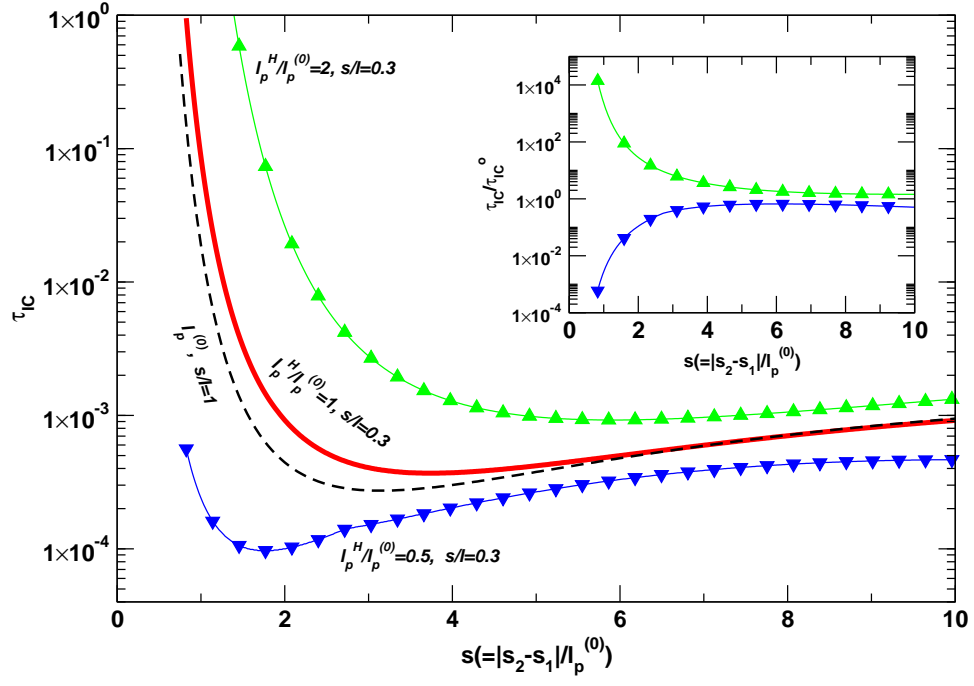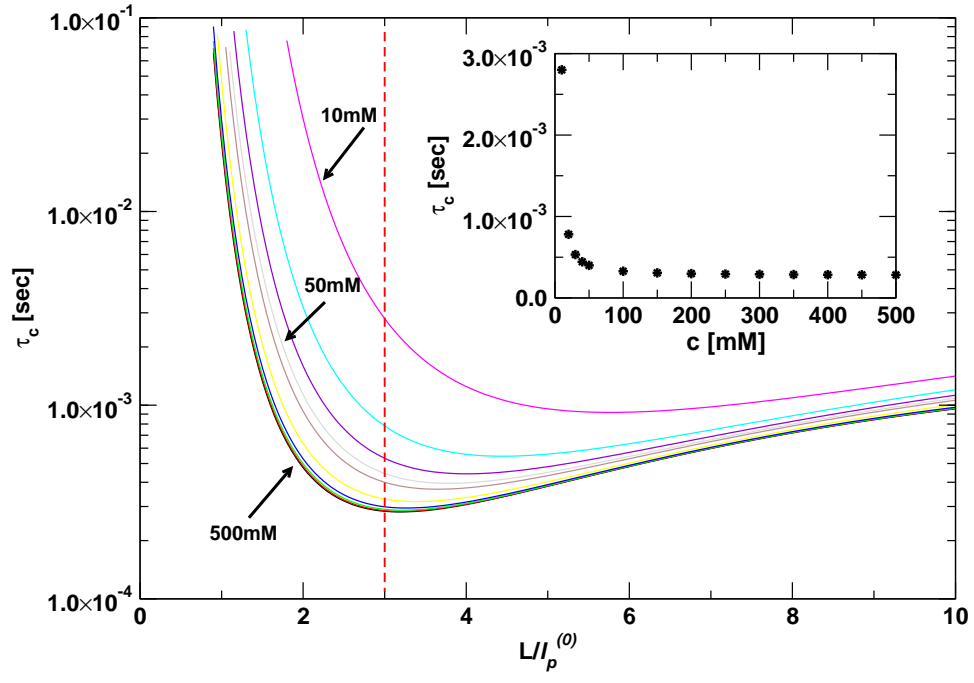

(b)

FIG. 1:

FIG. 2:

FIG. 3:

FIG. 4:

FIG. 5:

FIG. 6:

FIG. 7:

(a)

(b)

FIG. 8: