The thermodynamics of reversible cyclization in semiflexible polymers

Arti Dua and Binny J. Cherayil
Department of Inorganic and Physical Chemistry, Indian Institute of Science, Bangalore-560012, India

A recent model of the irreversible kinetics of ring formation in semiflexible polymers [J. Chem. Phys. 116, 399 (2002)] is generalized to the case of equilibrium cyclization, for which the rate constants for the forward and backward reaction are finite. The model is based on the diffusion-reaction formalism of Wilemski and Fixman [J. Chem. Phys. 60, 866 (1974)], and employs a path integral representation of the semiflexible chain (within a certain Gaussian approximation) to derive an expression for the steady state probability of occurrence of open configurations for given values of the chain length $N$, the reaction radius $a$, the degree of stiffness $z$, and the ratio of forward to backward reaction rates $k/k_r$. The steady state probability is used to calculate the free energy changes for the open-to-close transition. Chain rigidity is found to strongly influence the standard Gibbs free energy and enthalpy for the transition. While flexible chains tend to cyclize by virtue of their entropic elasticity alone, cyclization in semiflexible chains is also governed by the change in enthalpy between the open and closed states. The results are in qualitative agreement with the experimental measurements of Libchaber and co-workers.

I. INTRODUCTION

Long organic polymers are often satisfactorily modeled as completely flexible chains, but biological and other intrinsically semiflexible molecules typically require a more careful treatment of the effects of local rigidity. Several recent studies of biopolymers,1–4 in particular the work of Libchaber and co-workers on the cyclization of polynucleotides,5,6 highlight the inability of the fully flexible model of chain conformations to rationalize experimental data.

In the experiments of the Libchaber group, single stranded polydeoxyadenosine [poly(A)] and polydeoxythymidine [poly(T)] of between 8–30 residues are synthesized with a complementary 5-base sequence at either end, one sequence being terminated with a fluorophore and the other with a quencher. When dissolved in buffer solution, the polynucleotides cyclize through base pairing to form hairpin-loops that thermally fluctuate between open and closed states. These states are detected spectroscopically by fluorescence energy transfer, which varies as a function of the distance of separation between fluorophore and quencher. At large separations, the degree of fluorescence quenching is low, and at small separations, it is high. Open and closed states of the system exist in thermal equilibrium at long times, with low temperatures favoring closed states and high temperatures favoring open states. From the nature of the dynamic correlations between the fluctuations in the fluorescence yield, the free energy difference between open and closed conformations can be determined. The dynamics of poly(T) sequences are found to be governed solely by entropic elasticity, and fit the flexible coil model of polymers, but the dynamics of the poly(A) sequences are found to be additionally controlled by an enthalpic rigidity (presumably originating in base stacking), and cannot be fit to such a model.

The results illustrate the importance of chain rigidity in the equilibrium cyclization of short polymers.5–7 The dynamics of the forward reaction (i.e., the reaction leading irreversibly to loop structures) can be understood theoretically for chains that are assumed to be completely flexible. In particular, it is possible to successfully calculate the mean first passage time of closure for such chains as a function of parameters like the molecular weight of the chain and the distance of nearest approach of its reactive ends.8–15 We have recently shown that the calculation can be extended to treat semiflexible chains as well.16 However, a comparable treatment of the equilibrium process (which involves finite reaction rates for both the forward and backward directions) is presently lacking. In the context of the experiments by Libchaber and co-workers, we are therefore motivated to develop a model of equilibrium chain cyclization that can be used to determine the thermodynamics of the open-to-closed transition. We do this by extending the formalism of our earlier study of irreversible chain closure to incorporate the effects of the reverse reaction. This extended model also applies to chains of variable rigidity, from completely flexible to almost completely stiff. The effects of chain stiffness are accounted for by a single parameter, the persistence length, which is a measure of the degree of correlation between the tangent vectors at different points along the backbone of the chain. This is only a coarse-grained representation of the sequence-dependent rigidity of real polynucleotides, but it appears to capture the essential features of the actual system, and the predictions of the model are in satisfactory qualitative agreement with the results of experiment.

In the following section, we briefly recapitulate Fixman and Wilemski’s derivation of the equilibrium constant for the ring opening and closing reaction, starting from the Smolu-
A schematic drawing of the reversible cyclization of a long polymer with reactive end groups. $\tau_{\text{open}}$ and $\tau_{\text{close}}$ are, respectively, the average times of ring opening and ring closing.

The process of cyclization in the polynucleotide system of Ref. 5 involves an extended 5-residue binding region of complementary base pairs at the termini of the chains. The reaction we shall study is depicted schematically in Fig. 1, where this binding region is reduced to a pair of interacting points, with a reaction radius denoted by $a$. The replacement of the complementary base stem by twin reactive points is not expected to be a serious limitation, since experimentally it is the nucleation of the first pair of bases that is rate determining; thereafter, the binding of the remaining bases occurs far too quickly to be of much relevance to the present calculations.

The time evolution of open chain configurations is controlled by three factors: intramolecular diffusion, the intrinsic rate of binding of the reactive ends (once these ends are within the capture radius $a$), and the intrinsic rate of bond breakage in rings (once these rings are formed). The process is described by the following generalized Smoluchowski equation:

$$\frac{\partial \psi_i(\{r\},t)}{\partial t} = D_i \frac{\partial}{\partial r_i} \psi_i(\{r\},t) - k \psi_i(\{r\},t) + k_i \psi_i(\{r\},t),$$

where $D_i$ is the generalized diffusion operator for rings, and $k_i$ is the intrinsic rate constant for bond formation and bond cleavage.

The equilibrium constant for the interconversion of open and closed configurations is related to the ratio of their respective fractional concentrations at long times. These concentrations can be obtained from the long-time limit of the space integrals of the distribution functions $\psi_i(\{r\},t)$ and $\psi_i(\{r\},t)$, which in turn are obtained from the simultaneous solution of Eqs. (1) and (3).

$$\frac{\partial \psi_i(\{r\},t)}{\partial t} = D_i \frac{\partial}{\partial r_i} \psi_i(\{r\},t) + k_i(1 - p(t)) \psi_i^c(\{r\}),$$

where $D$ is the diffusion operator, and $k$ is the reaction constant, and $U$ is the reaction operator, and is defined in Eq. (2), with $D_0$ the diffusion constant, and $U$ the intersegment potential; $k$ and $k_i$ are the rate constants for bond formation and bond cleavage; $S$ is a source function that determines the reaction radius; and $\psi_i(\{r\},t)$ is the probability distribution function of the ring.

Once formed, rings also evolve in time according to a related Smoluchowski equation,

$$\frac{\partial \psi_i(\{r\},t)}{\partial t} = D_i \frac{\partial}{\partial r_i} \psi_i(\{r\},t) + k \psi_i(\{r\},t)$$

where $D_i$ is the generalized diffusion operator for rings, which differs from $D$ only in the nature of the intersegment potential. Its exact form will not be needed in these calculations.

The equilibrium constant for the interconversion of open and closed configurations is related to the ratio of their respective fractional concentrations at long times. These concentrations can be obtained from the long-time limit of the space integrals of the distribution functions $\psi_i(\{r\},t)$ and $\psi_i(\{r\},t)$, which in turn are obtained from the simultaneous solution of Eqs. (1) and (3). In practice, the solutions are difficult to obtain in closed form. As an approximation, the equilibrium constant $K$ is obtained as

$$K = \frac{1 - p}{p} = \frac{\tau_{\text{open}}}{\tau_{\text{close}}},$$

where $\tau_{\text{open}}$ and $\tau_{\text{close}}$ in this equation refer, respectively, to the mean times for chain opening and closing, and can be obtained from the fluorescence measurements described earlier. The second equality above therefore provides an experimental estimate of the equilibrium constant.

To solve Eq. (4), we introduce a Green’s function, $G(\{r\},\{r'\}|t,t')$, which solves the following equation:

$$\left(\frac{\partial}{\partial t} - D \right) G(\{r\},\{r'\}|t,t') = \delta(\{r\} - \{r'\}) \delta(t-t').$$

In terms of this Green’s function, Eq. (4) has the solution,
where ring and end effects in the ring polymer are small, the Hamiltonian of the first monomer located at the point can be approximated by the Hamiltonian

\[ \psi(\{r\},t) = \psi^{\text{eq}}(\{r\}) - k \int d\{r'\} \int_0^t dt' G(\{r\}\{r'\}|t,t') \times S(\{r'\}) \psi(\{r'\},t') + k_r \int d\{r'\} \times \int_0^t dt' G(\{r\}\{r'\}|t,t') \psi^{\text{eq}}(\{r'\})(1 - p(t')) . \]  

(7)

Equation (7) is a nonlinear integral equation for the distribution function \( \psi(\{r\},t) \) that does not appear to admit of a closed form solution. It is solved approximately, therefore, following the closure scheme of Wilemski and Fixman, which begins by multiplying Eq. (7) by the sink function \( S(\{r\}) \), and integrating over the position coordinates \( \{r\} \) to give

\[ \mu(t) = \mu_{\text{eq}} - k \int d\{r\} \int d\{r'\} \int_0^t dt' S(\{r\}) \times G(\{r\}\{r'\}|t,t') S(\{r'\}) \psi(\{r'\},t') + k_r \int d\{r\} \int d\{r'\} \int_0^t dt' S(\{r\}) G(\{r\}\{r'\}|t,t') \times \psi^{\text{eq}}(\{r'\})(1 - p(t')) , \]  

(8)

where

\[ \mu(t) = \int d\{r\} S(\{r\}) \psi(\{r\},t) . \]  

(9)

The equilibrium value of \( \mu(t) \), obtained by replacing \( \psi(\{r\},t) \) by \( \psi^{\text{eq}}(\{r\}) \) in Eq. (9), is the function \( \mu_{\text{eq}} \). The Wilemski–Fixman closure is the approximation

\[ \psi(\{r\},t) = \psi^{\text{eq}}(\{r\}) \frac{\mu(t)}{\mu_{\text{eq}}} . \]  

(10)

Now in general \( \psi^{\text{eq}}(\{r\}) \) is given by \( \delta(\{r_n\}) \times \exp(-\beta H_c(\{r\}))/Q_r \), where \( H_c \) is the Hamiltonian of the ring and \( Q_r \) its partition function, defined as

\[ Q_r = \int d\{r\} e^{-\beta H_c(\{r\})} \delta(\{r_n\}) . \]  

(11)

The delta function in these relations constrains the terminal monomer of the ring at the point \( r_n \) to occupy the site of the first monomer located at the point \( r_0 = 0 \). If it is assumed that end effects in the ring polymer are small, the Hamiltonian \( H_r \) can be approximated by the Hamiltonian \( H \) of a linear chain, so that one can also write

\[ \psi^{\text{eq}}(\{r\}) = \psi^{\text{eq}}(\{r\}) \frac{Q}{Q_r} \delta(\{r_n\}) . \]  

(12)

where \( Q \) is the partition functions for open conformations,

\[ Q = \int d\{r\} e^{-\beta H(\{r\})} . \]  

(13)

Furthermore, for a reaction in which cyclization involves only the terminal monomers of the chain, one of which is located at the fixed origin of coordinates, the sink function \( S(\{r\}) \) can be rewritten identically as

\[ S(\{r\}) = S(\{r_n\}) = \int d\{r\} \delta(\{r_n\}) S(\{r\}) . \]  

(14)

The above expression, when substituted into Eq. (8) along with Eqs. (10) and (12), leads to the following result:

\[ \mu(t) = \mu_{\text{eq}} - k \int_0^t dt' C(t-t') \mu(t') + k_r Q_r \int_0^t dt' D(t-t')(1 - p(t')) , \]  

(15)

where

\[ C(t-t') = \int d\{r\} \int d\{r'\} S(\{r\}) G(\{r\}\{r'\}|t,t') \psi^{\text{eq}}(\{r'\}) , \]  

(16)

\[ D(t-t') = \int d\{r\} \int d\{r'\} \delta(\{r_n\}) G(\{r\}\{0\}|t,t') \psi^{\text{eq}}(0) , \]  

(17)

with

\[ G(\{r\}\{r'\}|t,t') = \int d\{r\} \int d\{r'\} \delta(\{r_n\}) G(\{r\}\{r'\}|t,t') \]  

(18)

and

\[ \psi^{\text{eq}}(\{r\}) = \frac{1}{Q} \int d\{r\} e^{-\beta H(\{r\})} \delta(\{r_n\}). \]  

(19)

From Eq. (19) it follows that \( \psi^{\text{eq}}(0) \) is the ratio \( Q_r/Q \).

A further series of manipulations is carried out to derive an expression for the steady state probability \( p(s) \). The process starts by integrating both sides of Eq. (4) over the spatial positions of the monomers. Then, using the identity \( \int d\{r\} D\psi(\{r\},t) = 0 \) and the definition \( p(t) = \int d\{r\} \psi(\{r\},t) \) introduced earlier for the time-dependent probability, one easily derives the following expression for the Laplace transform \( \tilde{p}(s) \) of \( p(t) \), assuming an initially open configuration,

\[ \tilde{p}(s) = \frac{1}{s} - \frac{k \mu_{\text{eq}} \tilde{\omega}(s)}{(s + k_r)} , \]  

(20)

where \( \tilde{\omega}(s) \) is the Laplace transform of the function \( \omega(t) \), defined as the ratio \( \mu(t)/\mu_{\text{eq}} \). Equation (20), when substituted into the Laplace transform of Eq. (15), leads to an expression for the quantity \( \tilde{\omega}(s) \),

\[ \tilde{\omega}(s) = \frac{1}{1 + k \tilde{C}(s)/\mu_{\text{eq}} - k k_r \tilde{D}(s)/Q_r(s + k_r)} . \]  

(21)

It is more convenient at this stage to work with quantities

\[ H(t) = C(t) - \mu_{\text{eq}}^2 , \]  

(22)

\[ J(t) = D(t) - \mu_{\text{eq}} \psi^{\text{eq}}(0) , \]  

(23)
in which the infinite time limit of the functions $C(t)$ and $D(t)$ have been subtracted off. The Laplace transform of Eqs. (22) and (23), together with Eq. (21), when substituted into Eq. (20), produces

$$s \tilde{p}(s) = \frac{s + k_r + k(s + k_r) \tilde{H}(s)/\mu_{eq} - kk_r \tilde{Q}(s)/Q_r}{s + k_r + k \mu_{eq} + k(s + k_r) \tilde{H}(s)/\mu_{eq} - kk_r \tilde{Q}(s)/Q_r}.$$  

(24)

The steady state (long-time) probability can be obtained from Eq. (24) by taking the limit $s \to 0$. The result is

$$p_s = \frac{k_r}{k_r + k \mu_{eq} (1 + k \tilde{H}(0)/\mu_{eq} - k \tilde{Q}(0)/Q_r)}^{-1}$$

which is equivalent to

$$p_s = \frac{k_r}{k_r + k \mu_{eq} (1 + \mu_{eq} \int_0^\infty \frac{C(t)}{C(\infty)} - \frac{D(t)}{D(\infty)} dt)}^{-1}.$$  

(26)

This is one of the principal results of this section.

**B. Thermodynamic relations**

The steady state probability $p_s$ can be identified with the fraction of chains that exist in the open configuration in an imaginary ensemble of chains; the fraction of rings in this ensemble is therefore the quantity $1 - p_s$. Their ratio is the equilibrium constant $K$, which is given by

$$K = \frac{\tau_{\text{open}}}{\tau_{\text{close}}} = \frac{k \mu_{eq}}{k_r} \left[ 1 + k \mu_{eq} \int_0^\infty \frac{C(t)}{C(\infty)} - \frac{D(t)}{D(\infty)} dt \right]^{-1}.$$  

(27)

If instead of the various operations discussed above, Eq. (4) had simply been integrated over all coordinates, and then considered in the limit $t \to \infty$, the result would have been the steady state probability $p_s^{\text{eq}}$ of simple kinetic theory,

$$p_s^{\text{eq}} = \frac{k_r}{(k_r + k \mu_{eq})}.$$  

(28a)

The corresponding equilibrium constant $K^{\text{eq}}$ would have been given by

$$K^{\text{eq}} = \frac{k \mu_{eq}}{k_r}.$$  

(28b)

Equation (27) suggests that

$$\tau_{\text{close}} = k^{-1} \mu_{eq}^{-1} + \int_0^\infty \frac{C(t)}{C(\infty)} - \frac{D(t)}{D(\infty)} dt = \tau_{\text{close}}^{\text{kin}} + \tau_{\text{close}}^{\text{diff}}.$$  

(29)

In other words, the mean time of closing can be thought of as the sum of two distinct components, one associated with the time it takes for the two ends of the chain to react, once they are within reacting distance of each other, and the other associated with the time it takes for the process of molecular diffusion to bring the ends of the chain to this reaction distance. There is of course no diffusive component associated with $\tau_{\text{open}}$, so it can be identified with $k_r^{-1}$. Similarly, one can define a $K^{\text{diff}}$ as the ratio $\tau_{\text{open}}/\tau_{\text{close}}^{\text{diff}}$, and from Eq. (29), one then has

$$K^{\text{diff}} = \tau_{\text{open}} \left[ \int_0^\infty \frac{C(t)}{C(\infty)} - \frac{D(t)}{D(\infty)} dt \right]^{-1}.$$  

(30)

The overall equilibrium constant $K$ can therefore also be expressed as

$$K = \frac{K^{\text{eq}}}{(1 + K^{\text{eq}}/K^{\text{diff}})}.$$  

(31)

Furthermore, the equilibrium constant can be related to the standard Gibbs free energy for the reaction leading from open to closed configurations using

$$\Delta G = -RT \ln K.$$  

(32)

The above equation suggests that

$$K^{\text{diff}} = \tau_{\text{open}} e^{-\Delta G^{\text{diff}}/RT}$$  

(33)

which implies that

$$\tau_{\text{close}}^{\text{diff}} = \tau_{\text{open}} e^{-\Delta G^{\text{diff}}/RT}.$$  

(34)

In Ref. 5 the mean time of closing $\tau_{\text{close}}$ is assumed to be given by

$$\tau_{\text{close}}^{\text{diff}} = \tau_{\text{close}}^{\text{diff}}(\infty) e^{-\Delta H^{\text{diff}}/RT},$$  

(35)

where $\tau_{\text{close}}^{\text{diff}}(\infty)$ is the infinite temperature limit of $\tau_{\text{close}}$, and $\Delta H^{\text{diff}}$ is the standard enthalpy penalty for bringing the ends of the chain together through the process of diffusion. A direct comparison of Eqs. (34) and (35) leads to the relation

$$\tau_{\text{close}}^{\text{diff}}(\infty) = \tau_{\text{open}} e^{-\Delta S^{\text{diff}}/R},$$  

(36)

where $\Delta S^{\text{diff}}$ is the standard entropy change for the open to closed reaction. Equation (36) suggests that there is no energetic cost for bending a chain at infinite temperature. This in turn suggests that the infinite temperature limit of $\tau_{\text{close}}^{\text{diff}}$ corresponds to the closure time of a completely flexible chain. This would therefore imply that Eq. (36) is consistent with the standard Gaussian model of chain statistics, which contains contributions only from the entropic elasticity.

Equations (34) and (36) can be used together to estimate the diffusive contribution to the enthalpic penalty for chain deformation, which is given by

$$\frac{\Delta H^{\text{diff}}}{RT} = \ln \left( \frac{\tau_{\text{close}}}{\tau_{\text{close}}^{\text{diff}}(\infty)} \right).$$  

(37)

The thermodynamics of chain closure and its reverse is thus governed at the molecular level by the correlation functions $C(t)$ and $D(t)$. The following section discusses the calculation of these functions starting from the Green’s function $G(R,R',t,0)$, in conjunction with the sink function $S(\{R\})$, and the equilibrium distribution functions for open and closed chains. Details of the calculation have been discussed earlier in the context of irreversible chain closing, so the present treatment will be brief.
III. THE CORRELATION FUNCTIONS C(t) AND D(t)

Assuming that the end-to-end distance \( \mathbf{R} \) for both flexible and semiflexible chains is the sum of a large number of independent random displacements of the bond vector \( \mathbf{r}_i - \mathbf{r}_{i-1} \), the Green’s function \( G(\mathbf{R}, \mathbf{R}'; t, 0) \) can be expressed as a Gaussian distribution with a characteristic mean and standard deviation,\(^8,10,16\)

\[
G(\mathbf{R}, \mathbf{R}'; t, 0) = \left( \frac{3}{2\pi (R^2)_{eq}} \right)^{3/2} \frac{1}{(1 - \phi^2(t))^{3/2}} \times \exp \left[ -\frac{3(R - \phi(t)R')^2}{2(R^2)_{eq}(1 - \phi^2(t))} \right],
\]

(38)

where \( \phi(t) \) is the normalized time correlation function of the end-to-end distance. It is defined as

\[
\phi(t) = \frac{\langle \mathbf{R}(t) \cdot \mathbf{R}(0) \rangle_{eq}}{\langle R^2 \rangle_{eq}}.
\]

(39)

For a semiflexible polymer \( \langle R^2 \rangle_{eq} \) is given by\(^21\)

\[
\langle R^2 \rangle_{eq} = \frac{N}{p} \left[ 1 - \frac{1}{2pN} (1 - e^{-2pN}) \right].
\]

(40)

Here \( N = nL \) is the contour length of the chain, \( n \) being the number of monomers and \( L \) the persistence length,\(^22\) while \( p = l/N \) is the inverse of the persistence length. Although, in general, the distribution function of stiff chains is not correctly described by a Gaussian, a more accurate representation of the stiff chain limit would probably render the above calculations analytically intractable. The Gaussian assumption can be lifted, but at the expense of considerably more algebra.

We also assume that the sink function may be described by the Heaviside step function, i.e.,

\[
S(R) = \begin{cases} 
1; & R \leq a \\
0; & R > a,
\end{cases}
\]

(41)

where \( a \) is the minimum reaction distance between the ends of the chain.

Finally, we assume that the equilibrium end-to-end distribution function is also given by a Gaussian,

\[
\psi_{eq}(\mathbf{R}) = \left( \frac{3}{2\pi (R^2)_{eq}} \right)^{3/2} \exp \left[ -\frac{3\mathbf{R}^2}{2(R^2)_{eq}} \right]
\]

(42)

with the mean, \( \langle R^2 \rangle_{eq} \), given by Eq. (40).

Equations (38), (41), and (42) are now substituted into Eqs. (16) and (17). The integrand is then expanded as a series in the small parameter \( x_0 \), defined as

\[
x_0 = \frac{3a^2}{2(R^2)_{eq}}.
\]

(43)

Term by term integration of this series, followed by resummation, eventually leads to\(^5,16\)

\[
C(t) = \frac{16x_0^3}{9\pi} \left( 1 - \phi^2 + \frac{4}{5}x_0 \right)^{-3/2},
\]

(44)

\[
D(t) = \frac{4x_0^{3/2}}{3\pi^{1/2}} \left( 1 - \phi^2 + \frac{4}{5}x_0 \right)^{-3/4}.
\]

(45)

To calculate the steady state probability according to Eq. (26), the functions \( C(t) \) and \( D(t) \) in Eqs. (44) and (45) have now to be integrated over all time, after substituting for the function \( \phi(t) \). The calculation of \( \phi(t) \) is conveniently carried out in the continuum limit, where the set of discrete points \( r_i(t) \) defining the spatial positions of the \( n \) monomers at the time \( t \) is replaced by the set of continuous points \( r(s, t) \) at the same time, \( s \) being a real parameter that ranges from 0 to \( N \), the contour length of the chain (or from \(-N/2\) to \(+N/2\), if the midpoint of the chain is chosen to lie at the origin, as is done here). In this representation, and in units where the thermal energy \( k_BT = 1 \), the Hamiltonian \( H \) of a semiflexible polymer can be written as\(^23-26\)

\[
H = \nu \int_{-N/2}^{N/2} ds \frac{\partial r(s, t)}{\partial s} \frac{\partial r(s, t)}{\partial s} + \eta \int_{-N/2}^{N/2} ds \left[ \frac{\partial^2 r(s, t)}{\partial s^2} \right]^2 + v_0 (|u_{-N/2}|^2 + |u_{N/2}|^2),
\]

(46)

where

\[
\nu = \frac{3p}{2}, \quad \eta = \frac{3}{8p}, \quad v_0 = \frac{3}{4};
\]

(47)

and

\[
u = \frac{3p}{2}, \quad \eta = \frac{3}{8p}, \quad v_0 = \frac{3}{4};
\]

(47)

and

\[
\mathbf{u}(s, t) = \frac{\partial r(s, t)}{\partial s}.
\]

(48)

The first term in Eq. (46) describes the purely entropic connectivity of the chain, the second describes the energy cost involved in bending, and the last two terms act to suppress the excess fluctuations of the ends that lead to departures from the Kratky–Porod result for the end-to-end distance.\(^26\)

In writing \( H \) in this form, the geometric requirement that the tangent vector defined by Eq. (48) be unity everywhere is relaxed to the less stringent requirement that it be unity on average. In other words, the form of \( H \) above ensures that \( \langle |\mathbf{u}(s)|^2 \rangle = 1 \), where the angular brackets denote an equilibrium ensemble average over the conformations of the chain.

In continuum notation, the time correlation of the end-to-end distance is now given by

\[
\langle \mathbf{R}(t) \cdot \mathbf{R}(0) \rangle = \langle \mathbf{r}(N/2, t) \cdot \mathbf{r}(N/2, 0) \rangle \]

\[
+ \langle \mathbf{r}(N/2, t) \cdot \mathbf{r}(N/2, 0) \rangle
\]

\[
- \langle \mathbf{r}(N/2, t) \cdot \mathbf{r}(N/2, 0) \rangle
\]

\[
- \langle \mathbf{r}(N/2, t) \cdot \mathbf{r}(N/2, 0) \rangle.
\]

(49)

The position vectors \( r(s, t) \) evolve in time according to the equation\(^24,27\)

\[
\left[ \frac{\partial}{\partial t} + 2D_0 \eta \frac{\partial^3}{\partial s^3} - 2D_0 \nu \frac{\partial^2}{\partial s^2} \right] \mathbf{r}(s, t) = \mathbf{\theta}(s, t)
\]

(50)

and satisfy the boundary conditions

\[
\nu \frac{\partial \mathbf{r}(s, t)}{\partial s} + \eta \frac{\partial^2 \mathbf{r}(s, t)}{\partial s^2} \bigg|_{s=\pm N/2} = 0,
\]

(51)

\[
\nu_0 \frac{\partial \mathbf{r}(s, t)}{\partial s} + \eta_0 \frac{\partial^2 \mathbf{r}(s, t)}{\partial s^2} \bigg|_{s=\pm N/2} = 0,
\]

(52)

\[
\nu_0 \frac{\partial \mathbf{r}(s, t)}{\partial s} + \eta_0 \frac{\partial^2 \mathbf{r}(s, t)}{\partial s^2} \bigg|_{s=\pm N/2} = 0.
\]

(53)
In Eq. (50), \( \Theta(s,t) \) is a Gaussian random force with zero mean, whose second moment is given by
\[
\langle \Theta(s,t) \Theta(s',t') \rangle = \frac{2}{\zeta} \delta(s-s') \delta(t-t') I,
\] (54)
where \( I \) being the unit tensor, and \( \zeta \) the monomer friction coefficient.

Equation (50) is solved by
\[
r(s,t) = \int_{-N/2}^{N/2} ds_1 G_0(s,s_1|t) r(s_1)
+ \int_{-N/2}^{N/2} ds_1 \int_0^t dt_1 G_0(s,s_1|t-t_1) \theta(t_1) s_1 R(s_1, t_1),
\] (55)
where \( G_0(s,s_1|t) \) is the Green’s function of Eq. (50), which can be written as
\[
G_0(s,s_1|t) = \theta(t) \sum_{n=0}^\infty Q_n(s) \bar{Q}_n(s_1) e^{-D_0 \lambda_n t}. \] (56)

Here \( \theta(t) \) is the step function, and the \( Q_n(s), n = 0, 1, 2, ... \) are a complete orthonormal set of eigenfunctions that are the solutions of
\[
\left( 2 \eta \frac{\partial^4}{\partial s^4} + 2 v \frac{\partial^2}{\partial s^2} \right) Q_n(s) = \lambda_n Q_n(s) \] (57)
and that satisfy the same boundary conditions as defined by Eqs. (51)–(53). One can show that the \( Q_n(s) \) are of even or odd parity, and that they are given in general by
\[
Q_n(s) = \begin{cases} \left( \frac{C_n}{N} \right)^{1/2} \frac{\cos(\alpha_n s)}{\sin(\beta_n N/2)} & \text{even parity}, \\ + \left( \frac{\alpha_n}{\beta_n} \right)^2 \frac{\cosh(\alpha_n s)}{\sinh(\alpha_n N/2)} & \text{odd parity}, \end{cases} \] (58)
where \( C_n \) and \( D_n \) are normalization constants. Equation (57) together with the boundary conditions of Eqs. (51)–(53) also lead to a pair of eigenvalue equations. These are
\[
\beta_n^2 \sinh(\alpha_n N/2) \cos(\beta_n N/2) + \alpha_n^2 \cos(\alpha_n N/2) \sin(\beta_n N/2)
+ 2p(\alpha_n^2 + \beta_n^2) \sin(\beta N/2) \sinh(\alpha_n N/2) = 0,
\] (even parity),
\[
\beta_n^3 \cosh(\alpha_n N/2) \sin(\beta_n N/2) - \alpha_n^3 \sinh(\alpha_n N/2) \cos(\beta_n N/2)
- 2p(\alpha_n^2 + \beta_n^2) \cos(\beta N/2) \cosh(\alpha_n N/2) = 0,
\] (odd parity).

The parameters \( \alpha_n \) and \( \beta_n \) are related to each other and to the eigenvalues \( \lambda_n \) by the equations,
\[
\alpha_n^2 - \beta_n^2 = \frac{\nu}{\eta}
\] (62)
and
\[
\lambda_n = 2 \eta \alpha_n^2 \beta_n^2.
\] (63)

From Eqs. (55) and (56), it now follows that
\[
\langle r(s,t) \cdot r(s,0) \rangle = \sum_{n=0}^\infty \int_{-N/2}^{N/2} ds_1 Q_n(s) \bar{Q}_n(s_1) \langle r(s_1,0) \cdot r(s,0) \rangle
+ \langle r(s,0) \cdot r(s,0) \rangle \] (64)
with the equal time correlation function \( \langle r(s_1,0) \cdot r(s,0) \rangle \) given by
\[
\langle r(s_1,0) \cdot r(s,0) \rangle = \frac{\min[s,s_1]}{p} - \frac{1}{4p^2} [1 - (e^{-2ps} + e^{-2ps_1})] \] (65)
Substituting Eqs. (59), (64) and (65) into Eq. (49) and carrying out the integration over \( s_1 \), we obtain the following expression for the time correlation function of the end-to-end distance:
\[
\langle R(t) \cdot R(0) \rangle
= \frac{D_n}{N} \sum_{n=1}^\infty \left( \frac{2T_1 e^{-D_0 \lambda_n t}}{p} \right) \left( \frac{2T_2}{\beta_n^2} \right) \left( 1 + e^{-2pN} \right)
+ \frac{\beta_n^2 (1 - e^{-2pN})}{2p \alpha_n^2} \left( 1 - \frac{\alpha_n^4}{\beta_n^4} \right),
\] (66)
where
\[
T_1 = \tan(\beta_n N/2) + \frac{\alpha_n}{\beta_n} \tanh(\alpha_n N/2),
T_2 = \tanh(\alpha_n N/2),
T_3 = \tanh(\alpha_n N/2),
\] (67)
and
\[
\left( \frac{D_n}{N} \right) = \left( \frac{N}{2} \right) \left( \frac{\sec^2(\beta_n N/2) - \alpha_n^2}{\beta_n^2 \sech^2(\alpha_n N/2)} \right)
+ \left( \frac{3 \alpha_n^2 - \beta_n^2}{\alpha_n^2 + \beta_n^2} \right) T_2^{-1},
\] (68)
There are no contributions to \( \langle R(t) \cdot R(0) \rangle \) from \( n \) even.

The calculation of \( C(t) \) and \( D(t) \) now proceeds by first assigning definite values to the parameters \( p \) (the inverse persistence length), \( N \) (the contour length), \( a \) (the reaction radius), and \( t \) (the time). Once \( p \) and \( \nu \) are fixed, \( \alpha_n \) and \( \beta_n \) are found by numerically solving the simultaneous Eqs. (61) and (62) (for the odd parity solutions). Knowing \( \alpha_n \) and \( \beta_n \) the eigenvalue \( \lambda_n \) is calculated from Eq. (63), \( \alpha_n, \beta_n, \) and \( \lambda_n \) are then substituted into Eq. (66) along with a definite value for the diffusion coefficient \( D_0 \) (unity in all but one calculation, to be discussed later), and the result summed over all modes (in practice, no more than about 500), to yield a definite value for the correlation function \( \langle R(t) \cdot R(0) \rangle \) at the time \( t \). Normalizing this function by
\(\langle R^2 \rangle_{eq} \) yields \(\phi(t)\) at time \(t\). \(\phi(t)\) is then substituted into Eqs. (44) and (45) to produce \(C(t)\) and \(D(t)\), respectively, which are then numerically integrated over all time to produce the quantities needed in the evaluation of the equilibrium constant and other thermodynamic properties of the system. The entire procedure is repeated for different values of \(N, p, r\), and \(a\), thereby generating the thermodynamics of the system as a function of these parameters. In practice, the reaction radius \(a\) (expressed in units of \(l\), the persistence length) is fixed at the value 0.5, while the parameter \(p\) is specified by assigning a definite fixed value to a dimensionless stiffness parameter \(z\), defined as \(z = p N = N/l\).

IV. RESULTS AND DISCUSSION

As we have seen in Sec. II, the mean time of closing \(\tau_{\text{close}}\) of an initially open configuration of the polymer can be written as the sum of a time, \(\tau_{\text{diff}}\), associated with the diffusive process by which the ends of the polymer are brought to within a distance \(a\) of each other, and a time, \(\tau_{\text{kin}}\), associated with the intrinsic rate of reaction of these ends. One can therefore interpret the time scale of diffusion as the cyclization time of a reaction that takes place instantaneously whenever the ends of the chain are separated by the distance \(a\). Such a reaction is effectively irreversible, the rate constant for the forward direction being infinite. The calculations described in Ref. 16 (as well as those in Refs. 8 and 10 on which they are based) apply solely to this irreversible cyclization process. The present work generalizes these calculations to the case where both forward (ring closing) and backward (ring opening) reactions occur at finite rates and eventually establish an equilibrium between open and closed configurations. The nature of this equilibrium is under the macroscopic control of various thermodynamic potentials, which as we have now shown, can be expressed in terms of the microscopic parameters of the polymer cyclization model. The relation between these two levels of description is illustrated below in a series of graphs.

For instance, in Fig. 2, the full lines show the variation of the dimensionless standard free energy of cyclization, \(\Delta G\) [as calculated from Eqs. (27) and (32)] with the length \(N\) of the cyclizing chain for different fixed values of the stiffness parameter \(z\), at the fixed arbitrary value of 100 for the ratio \(k/k_r\), of the forward to backward reaction rates. \(\Delta G\) is seen, first of all, to vary linearly with \(N\), showing that the calculated thermodynamic potential is properly extensive in \(N\). Since larger values of \(z\) correspond to greater flexibility of the chain, \(\Delta G\) is also seen to increase as the chain becomes more and more rigid under conditions that tend to favor the forward reaction. As expected, it becomes harder for chains to cyclize as they become increasingly inflexible. Figure 2 quantifies this expected tendency thermodynamically. The dashed line in the figure is the variation of \(\Delta G\) with \(N\) at fixed \(l\) (as opposed to fixed \(z\)) and at fixed \(k/k_r = 100\) for the Rouse model,\(^{16}\) a flexible chain of coupled harmonic oscillators without excluded or hydrodynamic interactions. The free energy of cyclization for the Rouse model is estimated using Eqs. (44) and (45), with \(\phi(t)\) given by\(^{28}\)

\[
\phi(t) = 8 \sum_{n \text{ odd}} \frac{1}{n^2} e^{-3n^2 \pi^2 D \sigma t / N^2}.
\]

For this model, by virtue of the fact that \(z\) increases as \(N\) increases at fixed \(l\), the free energy starts out at roughly the same \(\Delta G\) value as our continuum chain model with \(z = 50\), but then increases more slowly with \(N\) than such a chain. The variation reflects the crossover from less to more flexible chain behavior as the polymer becomes longer.

Figure 3 shows the variation of \(\Delta G\) with \(N\) at three different \(k/k_r\) ratios (100:1, 1:1, and 1:100) for a chain that is relatively stiff (\(z = 1.0\)). The lowest curve in the figure is identical to the topmost curve of Fig. 2. The other two curves in Fig. 3, corresponding to increasingly smaller values of the
ratio $k/k_r$, reflect the greater free energy costs that have to be paid to cyclize chains under conditions where thermal fluctuations favor open configurations.

Figures 4 and 5 are, respectively, plots of the dependence on $N$ of the diffusive components of the equilibrium constant and the standard enthalpy of cyclization at different $z$ and at fixed $k/k_r = 100:1$, as calculated from Eqs. (30), (37), (44), and (45). The enthalpy change varies in much the same way as the free energy, the enthalpy penalty being greater for the cyclization of chains with smaller $z$. In the flexible limit ($z \geq 50$), Figs. 2 and 5 suggest that chain closure is governed principally by entropic elasticity, rather than by enthalpy. This is in qualitative agreement with data from the experiments of Libchaber and co-workers. However, the predicted slopes of the $\Delta H$ versus $N$ curves, which in our model are the same for all $z$, differ from the experimental slopes, which are greater for chains of larger persistence length (or sequence dependent rigidity).

In general, experimental enthalpies are determined by measuring the variation of the cyclization time with temperature $T$. In our model, temperature enters essentially through the parameter $D_0$, the monomer diffusion coefficient (strictly speaking, the persistence length is also temperature dependent, but in principle it can be held fixed as temperature varies), so it is possible to study the role of $T$ in our model by varying $t_{\text{close}}^{\text{diff}}$ with $D_0$. The nature of this variation is shown in Fig. 6, the abscissa being plotted in terms of $1/D_0$ rather than $D_0$ itself; $1/D_0$ is directly related to the viscosity. The predicted linear dependence of $t_{\text{close}}^{\text{diff}}$ on $1/D_0$ is in qualitative agreement with the results of recent experiments carried out by Wallace et al.

Some of the discrepancies that we observe between the calculated and observed thermodynamics of cyclization can probably be traced to limitations in our representation of the very stiff regime. Our calculations presently apply only to chains of low to moderate rigidity, ($z \gg 1$), as the solutions to the various equations that produce the integrated values of the correlation functions $C(t)$ and $D(t)$ become numerically unstable when $z < 1$. This loss of numerical stability may in turn be due to the assumed Gaussian form for the end-to-end distribution functions of the polymer. The assumption is likely to break down as the chain becomes increasingly stiff. Although the deficiency can be corrected in principle by working with non-Gaussian distributions in the stiff limit, the attendant algebra is expected to be very unwieldy. At the same time, since the vertical stacking interactions responsible for the sequence dependent rigidity in real DNA (Ref. 30) are only crudely accounted for in our model by the single parameter $z$, it may be possible to improve the model’s predictive ability by treating these interactions more realistically.
20. The equation satisfied by the Green’s function in Ref. 16 [Eq. (A2)] contains a typographical error: it should be identical to the equation [Eq. (6)] satisfied by the Green’s function in this paper.
22. It is somewhat more common to define the persistence length as $l/2$, rather than as $l$, as we have done here. The present convention has been adopted purely as a matter of convenience.
27. In our earlier paper, Ref. 16, the expression for the Lagrangian $\mathcal{L}$ that is used to derive the evolution equation of the monomer position should have read $\mathcal{L} = (1/2m)\int ds \dot{p}(s)^2 - H$, where $m$ is the mass of the monomer and $p(s)$ its momentum.