Metastable Tight Knots in Semiflexible Chains

Liang Dai, † C. Benjamin Renner, ‡∥ and Patrick S. Doyle †‡∥

‡ BioSystems and Micromechanics IRG, Singapore—MIT Alliance for Research and Technology Centre, Singapore 117543, Singapore
∥ Department of Chemical Engineering, Massachusetts Institute of Technology (MIT), Cambridge, Massachusetts 02139, United States

ABSTRACT: Knotted structures can spontaneously occur in polymers such as DNA and proteins, and the formation of knots affects biological functions, mechanical strength and rheological properties. In this work, we calculate the equilibrium size distribution of trefoil knots in linear DNA using off-lattice simulations. We observe metastable knots on DNA, as predicted by Grosberg and Rabin. Furthermore, we extend their theory to incorporate the finite width of chains and show an agreement between our simulations and the modified theory for real chains. Our results suggest localized knots spontaneously occur in long DNA and the contour length in the knot ranges from 600 to 1800 nm.

1. INTRODUCTION

Knotted structures can occur in polymers, such as DNA and proteins. The investigation of knots in mathematics has a rich history, and recently research into these structures is increasingly motivated by biological relevance and practical concerns such as DNA sequencing. Much like the well-known tangling of jumbled strings, linear polymers can spontaneously form knotted structures, and their topologies can be determined via procedural closing schemes. Knots (most frequently simple ones) are present in the DNA contained in viral capsids and increasing increasingly motivated by biological relevance and practical concerns such as DNA sequencing. Much like the well-known tangling of jumbled strings, linear polymers can spontaneously form knotted structures, and their topologies can be determined via procedural closing schemes. Knots (most frequently simple ones) are present in the DNA contained in viral capsids and influence the rate of ejection of the viral genome, and a number of knotted protein structures have been identified. Beyond their biological importance, knots in polymers can affect their mechanical strength, rheological properties, and may play a role in crystallization, cause jamming during nanopore DNA sequencing, or be used for controlled drug delivery.

For these reasons, there is a desire to elucidate the size and probabilities of knots in polymer molecules. Accordingly, simulations addressing this issue have been performed for numerous cases: linear, and circular chains, ideal and self-avoiding chains, flexible and semiflexible chains, lattice and off-lattice models, good and bad solvents, as well as in free space, confinement, and under tension. An intriguing finding from simulations is that the cores of knots very often localize at small portion of chain. For example, Katritch et al. found the most probable size of trefoil knot on an unconfined circular ideal chain is only seven segments. The localization of polymer knots has also observed in experiments, and several theories have been developed to explain this behavior. The theory of Grosberg and Rabin employed the idea, similar to the tube theory for polymer melts, that the chain in the area of the knot forms a self-consistent confining tube. The bending energy and confinement free energy within a knot tend to swell and shrink the knot respectively, leading to a localized, metastable knot that is likely to untie by diffusing to the end of the molecule. Beyond qualitatively predicting knot localization, this theory is able to predict the size of such a metastable knot in a polymer, but the theory has yet to be rigorously tested against simulation or experiment.

In the current study, we validate the Grosberg–Rabin theory for metastable knots with computer simulations of long, linear wormlike polymers. We extend the theory to incorporate the finite thickness of polymer chains and similarly validate it with computer simulations. Our results demonstrate a quantitative agreement between simulation and theory and provide testable predictions for sizes of knots in dsDNA, which has been used in previous experimental work on knotting.

2. THEORY AND SIMULATION

2.1. Theory of Knots in Semiflexible Chains. We first recall the Grosberg–Rabin theory for a knot wormlike polymer of zero thickness. The central tenet of this theory is that the polymer contour in the knotted region is self-confined by a virtual tube (Figure 1). The tube can be imagined as a tight knot formed by pulling both ends of a rope with diameter $D$. The free energy cost to form a knot contains contributions due to both bending and confining the polymer contour in the knotted region. First, the bending energy $\mathcal{F}_{\text{bend}}$ in the knot core...
scales as \( L_{\text{knot}} R_{\text{knot}}^{-2} \), where \( L_{\text{knot}} \) and \( R_{\text{knot}} \) are the contour length and the size of the knot core, respectively. Second, the free energy of confining the polymer within the knot core is given as \( F_{\text{conf}} \sim L_{\text{knot}} D^{-2/3} \left( L_p^{-1/3} \right)^2 \) after applying the Odijk scaling,31 where \( L_p \) is the persistence length. For a tight knot, the quantities \( L_{\text{knot}} \) and \( D \) are proportional to each other. Accordingly, we replace \( R_{\text{knot}} \) and \( D \) with \( L_{\text{knot}} \) after ignoring numerical coefficients. Note that \( L_{\text{knot}} \) can be measured in simulations, while \( D \) is difficult to determine directly. Accordingly, Grosberg and Rabin obtain the total free energy \( F_{\text{slc}} = F_{\text{conf}} + F_{\text{bend}} \)

\[
\frac{F_{\text{slc}}}{k_B T} = k_1 \left( \frac{L_{\text{knot}}}{L_p} \right)^{-1} + k_2 \left( \frac{L_{\text{knot}}}{L_p} \right)^{1/3}
\]

(1)

where \( k_1 T \) is the thermal energy, and \( k_1 \) and \( k_2 \) are prefactors that take account of all numerical coefficients ignored in the derivation. The first and second terms in the above equation tend to swell and shrink the knot size, respectively. The competition of these terms leads to a local minimum of free energy, which corresponds to a metastable knot.

Next, we adapt the Grosberg-Rabin theory to real chains with finite thicknesses. For a chain with an effective width \( w \) confined in a tube with a diameter \( D \), the effective diameter of the confining tube becomes \( D_{\text{eff}} = D - w \) due to the repulsion between the chain and tube walls. As a result, the confinement free energy becomes \( F_{\text{conf}} \sim L_{\text{knot}} (D - w)^{-2/3} L_p^{-1/3} \). Similar to the previous study,28 we define \( p \) as the ratio of \( L_{\text{ideal}} \) to \( D \)

\[
p \equiv L_{\text{ideal}} / D \approx L_{\text{knot}} / D
\]

(2)

where \( L_{\text{ideal}} \) is a topological property defined for a maximally inflated knot.32

We then obtain \( F_{\text{conf}} \sim L_{\text{knot}} (pD - pw)^{-2/3} L_p^{-1/3} = L_{\text{knot}} (L_{\text{knot}} - pw)^{-2/3} L_p^{-1/3} \). The free energy cost of forming a knot on a real chain becomes

\[
\frac{F_{\text{conf}}}{k_B T} = k_1 \left( \frac{L_{\text{knot}}}{L_p} \right)^{-1} + k_2 L_{\text{knot}} (L_{\text{knot}} - pw)^{-2/3} L_p^{-1/3}
\]

(3)

It is easy to see that when \( w = 0 \), eq 3 returns to eq 1. The size of metastable knots for real chains can be calculated by minimizing \( F \) with respect to \( L_{\text{knot}} \). This calculation requires the values of \( p \), \( k_1 \), and \( k_2 \). We estimate these values for trefoil knots as follows. The value of \( p \) is estimated to be 12.4, which is calculated for the tightest trefoil knot with both ends in a line.32 To estimate \( k_1 \), we assume the bending in the core of trefoil knots is uniform and the total bending angle is \( \theta_{\text{total}} \). Then, the bending energy follows \( (1/2) \theta_{\text{total}}^2 / L_p / L_{\text{knot}} \). To form any knot, the total bending angle should be larger than \( 2 \pi \), so \( k_1 \) should be larger than 19.7. To estimate \( k_2 \), we apply the scaling of confinement free energy with a prefactor \( 3^{1/3} \approx 2.4 \) to the estimation of \( k_2 \approx 2.4 \times (D - w)^{-2/3} L_p^{-1/3} \), leading to an estimate of \( k_2 \approx 12.9 \). We will show the estimations of \( k_1 \) and \( k_2 \) are close to the fitted values from simulation results while the estimation of \( k_2 \) is much larger than the fitted value.

2.2. Simulations of Knots in Semiflexible Chains. To examine the theory, we performed simulations of long open semiflexible chains and analyzed the knotted conformations. The polymer chain is modeled as a string of touching beads.34 The diameter of each bead equals the effective chain width \( w \), and the contour length \( L \) is thus \( L = (N - 1) w \), where \( N \) is the number of beads. There are only two interactions between beads: the pairwise hardcore repulsion between beads and the bending energy \( F_{\text{bend}}(\theta) / k_B T = (1/2) L_p / w \theta^2 \), with bending angles \( \theta \), to reproduce the persistence length \( L_p \). We use the Pruned-enriched Rosenbluth method (PERM) algorithm35 to generate polymer configurations and then analyze the topologies of these configurations after closing the both ends. In the following paragraphs, we will describe how we implement PERM algorithm, how we close the both ends of a linear chain, and how we determine the size of knot core if the chain is knotted.

The PERM algorithm generates chain configurations by a growth process.35,36 The growth starts from a bead at the origin (0,0,0). In each step of growth, a new bead (the i-th bead) is placed at the end of the chain. The orientation of the newly added bead follows a Boltzmann distribution of bending energies \( \exp[-E_{\text{bend}}(\theta)/k_B T] \), where \( \theta \) is the bending angle formed by the \((i - 2)\text{th}, (i - 1)\text{th}, \) and \( i\text{th} \) beads. If the new bead overlaps with any other bead, then this chain dies. In the simulation, each chain is grown in a batch of \( N_c \) chains. After each step of adding a bead, a few chains may die. The surviving chains are duplicated, which is the so-called enrichment. Without enrichment, more and more chains die as the length of chain grows, and we cannot obtain enough sampling of long chains. The idea of the PERM algorithm is to duplicate the surviving chains and reduce the statistical weight of these surviving chains such that these chains are not overrepresented in the final samplings. We implement the enrichment as follows. In each simulation, we grow a batch of \( N_c \) chains. The initial weights of these chains are \( W_{ij} = 1 / N_c \), for \( i = 1 \), where \( i \) denotes the chain length and \( j \) denotes the index of chain. Suppose \( N_j \) chains die after adding the \( i \text{-th} \) bead. Then, we randomly pick \( N_j \) chains from \((N_c - N_j) \) survived chains and duplicate these \( N_j \) chains so that the number of chains remains at \( N_c \) during chain growth. Considering that all surviving chains have the same probability \( q = N_j / (N_c - N_j) \) to be duplicated, the weights of all chains are reduced by a factor \( 1/(1 + q) = (N_c - N_j)/N_c \). That is, \( W_{ij+1} = W_{ij} (N_c - N_j) / N_c \) after adding \((i+1)\text{-th} \) bead to the \( j \text{-th} \) chain. When the chain length (number of beads) reaches the desired chain length, the growth stops, and the configurations of \((N_c - N_j) \) survived chains and the associated weights \( W_{ij} \) are used to calculate the probability of the trefoil knot \( f = \left( \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} A(j) W_{ij} \right) / \left( \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} W_{ij} \right) \), where \( A(j) \) equals one when the \( j \text{-th} \) chain is a trefoil knot, and equals zero when the \( j \text{-th} \) chain is not a trefoil knot.

To determine the topology of an open chain, the chain must first be closed. In the current study, we employ the minimally interfering closure scheme.3 After closing an open chain, we calculate the Alexander polynomial to identify the topology.

To identify the core of a trefoil knot, we cut the maximum number of beads from each end of the chain while the trefoil knot remains. The two boundaries of knot core are determined sequentially. First, we keep one end unchanged and cut beads from the other end to determine the boundary. After obtaining this boundary, we cut beads to determine the other boundary.

Figure 1. Illustration of a trefoil knot on an open chain (red). The subchain with a contour length of \( L_{\text{knot}} \) in the knot core is confined in a virtual tube (gray) with a diameter of \( D \).
We find that the size of knot core may depend on which boundary is determined first. For consistency, we calculated the knot size both ways and used the smaller value as the final size of knot core because the core of knot corresponds to the smallest portion of the chain that forms the knot.

3. RESULTS AND DISCUSSIONS

Figure 2a shows the probability of forming a trefoil knot as a function of the rescaled knot size for a wormlike chain of zero thickness. The contour length is \( L = 400L_p \). The probability \( f(x) \) is normalized such that \( \int_0^{L/k} f(x) \, dx = f_{total} \), where \( x = L_{knot}/L_p \) and \( f_{total} \approx 0.23 \) is the total probability of trefoil knot. The probability exhibits a peak at a certain knot size \( L_p \approx 12L_p \). The value \( L_{knot} \) is insensitive to increasing the contour length (see the Supporting Information) since the knots are localized. The most probable size of a trefoil knot in a previous simulation of a circular freely jointed ideal chain by Katritch et al.\(^{16}\) is seven segments, each corresponding to \( 2L_p \). Thus, their finding of a knot size of \( 14L_p \) is close to the \( 12L_p \) reported here.

In Figure 2b, the probability is converted to the potential of mean force (PMF) using \( F = \ln(f) - F_0 \) where \( F_0 = \min(-\ln(f)) \) to offset the minimum. The PMF curve exhibits a local minimum in a potential well, corresponding to a metastable knot. This minimum is a local minimum because the global minimum corresponds to the unknotted state. As can be seen from our simulation results, the potential well is relatively broad, and the depth is \( \sim 1 \, k_BT \) up to \( L_{knot}/L_p \approx 50 \), which is over 3 times the size of the knot with the minimum free energy. This observation indicates the metastable minimum is relatively broad, and long chains (several times \( pL_p \)—a characteristic size of a tight knot) are required to observe it. In the simulations of knots by Zheng and Vologodskii,\(^{37}\) shallow or even flat potential wells are observed for the \( 7_1 \) and \( 10_{1,2,3} \) knots (both much larger than the trefoil knot) under a slight stretching force. The authors concluded no metastable knot size exists for these topologies, attributing the slight minimum they observe to the applied stretching force. The apparent difference between their results and the present work may arise due to the very shallow nature of the potential well—if a metastable knot size exists for these large knots, extremely long chains (several multiples of \( pL_p \)) would be needed to visualize the well. For the long chains in our simulations, however, the depth of the potential well we observe is \( \sim 6 \, k_BT \) for \( L_{knot}/L_p = 400 \), which suggests that the spontaneous unknotting process of a knot on a long and thin chain is more likely through the diffusion of knot along the chain than through the swelling of knot toward the full contour length.\(^{38,39}\)

We use eq 1 to fit the PMF curve in the range \( L_{knot}/L_p \leq 28 \) and determine the prefactors \( k_1 = 17.06 \) and \( k_2 = 1.86 \). The fitted line (red) deviates from simulation results for \( L_{knot}/L_p > 30 \). This is expected from the theory because the theory assumes the chain is strongly confined in a virtual tube, and the confinement free energy follows the Odijk scaling \( F_{conf} \sim D^{-2/3} \), where \( D \) is the diameter of the virtual tube. This assumption becomes invalid when \( D \equiv L_{knot}/p > 2L_p \). If we use the estimation \( p = 12.4 \) as discussed above, then the Grosberg-Rabin theory is supposed to be inapplicable for \( L_{knot} > 2pL_p = 24.8L_p \) which is consistent with the results in Figure 2b. The assumption of Odijk-like confinement, however, is always valid for small knots. As a result, the fit of the theory matches the simulation data well through the smallest observed knots, \( L_{knot} = 2L_p \).

The fitted value of \( k_1 = 17.06 \) is slightly less than the estimate for the lower bound of 19.7, as described above. This discrepancy likely arises from the fact that the bending energy in the unknotted subchains is ignored when we estimate the total bending energy cost of forming a knot. To re-examine the bending energy term, \( 17.06L_p/L_{knot} \) we directly analyzed the increase in bending energy due to forming a knot from the chain configurations in simulations. This analysis confirms that the term \( 17.06L_p/L_{knot} \) indeed roughly captures the bending energy increase in knots (see the Supporting Information). On the other hand, for the confinement free energy, the fitted value of \( k_2 = 1.86 \) is much less than the estimate of \( k_2 = 12.9 \). This deviation may result from the assumption that the virtual tube of the knot (Figure 1) has hard, solid walls. In reality, the walls of the virtual tube are made of fluctuating sections of the same chain, and this “softer” confining tube results in the smaller prefactor from the simulation data.

Figure 3a shows the probability of forming a trefoil knot as a function of the rescaled knot size for real chains with different chain widths. As the chain width increases, the total probability of forming a trefoil knot monotonically decreases. The most probable knot size, \( L_{knot}^* \) also increases with increasing chain width, as expected. We analyze the most probable knot size \( L_{knot}^* \) rather than the mean knot size \( \langle L_{knot} \rangle \) because \( L_{knot}^* \) is insensitive to the contour length and is a local quantity (see the Supporting Information). On the other hand, the mean size \( \langle L_{knot} \rangle \) strongly depends on the contour length,\(^{6,15,40}\) due to a long tail in the distribution of knot sizes. The inset of Figure 3a shows the total probability of forming a trefoil knot as a function of the rescaled chain width (circle symbols). The probability decreases from 22.8% to 0.57% as \( w/L_p \) increases from 0 to 0.5. We also plot the total probability (square symbols) for \( L_{knot}/L_p \leq 100 \) because some natural ambiguity arises in classifying knots on open chains when the size of the knot approaches the total contour length. Using this criterion, the probability decreases from 18.8% to 0.33%. Furthermore, this only modest reduction in probability when the size of the knot is restricted to a small fraction of the total contour underscores the notion that these knots are localized.

---

**Figure 2.** (a) Probability of a wormlike chain containing a trefoil knot with \( L = 400L_p \). (b) Potential of mean force as a function of knot size. The line of best fit is shown in red: \( y = 17.06x^{-1} + 1.86x^{1/3} - 5.69 \). Both insets show curves over wider ranges.
The depth of potential well is larger than 1 small, and we simulate results and theoretical predictions when 100. When 1 \leq \mu \leq 0.1, the depths further increase (See the Supporting Information). We note that the potential well becomes shallower as the chain width increases, indicating that knots on thick chains, while less probable, will exhibit much more variation in size.

The dependence of L_{knot} on w is plotted in Figure 4 and compared with theoretical predictions from eq 3. The values of k_1 and k_2 are already determined from the fit in Figure 2 for the wormlike chain of zero thickness, so only p is used as a fitting parameter. Recall that eq 3 is supposed to be applicable only when the knot is small. So we manually tune p to match simulation results and theoretical predictions when L_{knot} is small, and we find p = 16 is suitable for w/L_p \leq 0.25. Furthermore, p = 16 is a reasonable value for the ratio of L_{knot} to the tube diameter D, considering that p is estimated as 12.4 for the tightest trefoil knot, which is the lower bound of p value. For large knots, the Odijk scaling F \sim D^{-2/3} overestimates the confinement free energy. Recall that the overestimation in confinement free energy tends to shrink the knot, and leads to the observed discrepancy between the simulation results and the theoretical prediction for large knots on thick chains. Similar analysis was performed for the figure eight (4_1) knot (see the Supporting Information), yielding a fitted value of p = 38 for 4_1 knot.

After obtaining the fitted value of p = 16 for the trefoil knot, we calculate the potentials of mean force using eq 3 for w = 0.1L_p and w = 0.2L_p, and compare with the ones obtained from simulations, as shown in Figure 5. Near the minima, the theoretical predictions agree with the simulation results.

From the metastable knot sizes in Figure 4 and the fitted value of p = 16, we can infer that the diameter of the virtual tube D = L_{knot}/p varies from 0.75L_p to 2.2L_p as the chain width increases from 0 to 0.5L_p. A similar virtual tube also exists in entangled polymers, and the effective potential in the virtual tube has been calculated by Zhou and Larson. If we set 1 \mu T as a cutoff in the tube potential, the tube diameter is roughly 8 L_p, while the effective chain width w roughly equals L_p in their calculation. An important difference in these scenarios is that in the case of a knot, one chain is confined by itself, while in the case of entanglement, one chain is usually confined by other chains.

We now turn our attention to “real” polymers of nonzero thickness. The total probability of forming a trefoil knot as a function of the rescaled chain width is shown in the inset of Figure 3a. The simulation parameters we employed in this study are directly comparable to \lambda-DNA, which is widely used in experiments. The contour length in simulations is 400 L_p which is close to the case of \lambda-DNA, considering that the YOYO-1 intercalated \lambda-DNA has a contour length about 22 \mu m and a persistence length about 50 nm. The range of chain widths in our simulations is from 0 to 0.5L_p which is within the range of the effective width of DNA at experimental ionic strengths. These simulations predict the knotting probability of \lambda-DNA can be around 10% for high salt conditions. The contour length of the metastable trefoil knot typically range from 600 to 1800 nm, while the knot radius of gyration ranges from 65 to 138 nm (see the Supporting Information). Note that the knotting probability of a circular real DNA, much shorter than 48.5-kb \lambda-DNA, has been shown experimentally to be 4% in high salt solution.
As discussed above, the modified Grosberg-Rabin theory is not applicable for larger knots because the Odijk scaling becomes invalid when $D \gg L_p$. For these more swollen knots, the virtual tube is wide, and the confinement free energy is proportional to the number of blobs $N_{blob} = L_{knot}/D = p$ as discussed by Grosberg and Rabin. In this case, the confinement free energy is independent of $L_{knot}$ and thus no longer tends to shrink the knot core. Accordingly, Grosberg and Rabin suggest that the force driving the chain toward tight knots disappears when the tube is wider than $L_p$. However, our simulation results in Figure 3a show that the PMF curve keeps increasing with the increasing size of knot core. Other factors may need to be considered for the spontaneous tightening of large knots.

**CONCLUSIONS**

We have extended the theory of Grosberg and Rabin to calculate the size and distribution of knots on a real, semiflexible polymer molecule and validated the predictions of this theory with computer simulations. These tight metastable knots exist on semiflexible chains due to the competition of bending energy and confinement free energy, which tend to swell and shrink knot size, respectively. The metastable knots are trapped by potential wells with depth of several $k_B T$. Our simulation results should be universal because the model contains only two parameters: $L_p$ and $w$, or a single dimensionless parameter $w/L_p$. Looking forward, we expect that the Grosberg–Rabin theory can be further modified for chains confined in channels. Finally, our simulations are directly comparable to λ-DNA at experimentally relevant ionic strengths, and we hope that future experiments will allow for further testing of these ideas.

**ASSOCIATED CONTENT**

 Supporting Information
Dependence of results on bond and contour length, distributions of knot positions, results for the $4j$ knot, bending energy analysis within knots, and radii of gyration of knots. This material is available free of charge via the Internet at http://pubs.acs.org/.

**AUTHOR INFORMATION**

*(P.S.D.) E-mail: pdoyle@mit.edu.

**Author Contributions**

#These authors contributed equally.

**Notes**

The authors declare no competing financial interest.

**ACKNOWLEDGMENTS**

This research was supported by the National Research Foundation Singapore through the Singapore MIT Alliance for Research and Technology’s research program in BioSystems and Micromechanics, the National Science Foundation (Grant No. 1335938). The authors thank the center for computational science and engineering in National University of Singapore for providing the computational resources.

**REFERENCES**