Role of Helicity in DNA Hairpin Folding Dynamics

Huaping Li and Alkan Kabakçıoğlu*
Department of Physics, Koç University, Istanbul, 34450, Turkey

(Received 16 May 2018; published 25 September 2018)

We study hairpin folding dynamics by means of extensive molecular dynamics simulations, with particular attention paid to the influence of helicity on the folding time. We find that the dynamical exponent α in the anomalous scaling $n(t) \sim t^{1/\alpha}$ of the hairpin length n with time changes from 1.6 ($\simeq 1+\nu$, where ν is the Flory exponent) to 1.2 ($\simeq 2\nu$) in three dimensions, when duplex helicity is removed. The relation $\alpha = 2\nu$ in rotationless hairpin folding is further verified in two dimensions ($\nu = 0.75$) and for a ghost chain ($\nu = 0.5$). Our findings suggest that the folding dynamics in long helical chains is governed by the duplex dynamics, contrasting the earlier understanding based on the stem-flower picture of unpaired segments. We propose a scaling argument for $\alpha = 1+\nu$ in helical chains, assuming that duplex relaxation required for orientational positioning of the next pair of bases is the rate-limiting process.

DOI: 10.1103/PhysRevLett.121.138101

DNA/RNA hairpin folding is the temperature-driven self-assembly of a palindromic nucleic acid composed of two complementary sequences linked by a relatively short "loop" segment. Transcription and folding of small hairpins (e.g., siRNAs, miRNAs) help initiate biochemical reactions, cell signaling, gene expression, and a viral response in many organisms [1–4]. Their synthetic counterparts are used ubiquitously in biotechnological applications, such as in CRISPR [5–8]. Interest in the folding dynamics of such molecules has grown recently due to the availability of new experimental techniques that allow high-resolution observations both in time and space [9–11]. As revealed by numerical simulations and experiments, the formation of a folding nucleus at the center of the hairpin is the time-limiting step [12–16]. Nonetheless, actual folding time (zippering after nucleus formation, sometimes referred as the transition-path time) has been the focus of several recent studies due to its anomalous character [17–19].

Progression of zippering can be monitored through the duplex length, $n(t)$, which serves as the natural reaction coordinate. Earliest theoretical models for predicting the folding time, such as the zipper model, were based on the equilibrium free-energy difference between paired (double-strand) and unpaired (single-strand) states. Such considerations predict a ballistic process $n(t) \sim t$ for $T < T_c$ and a diffusive one for $T = T_c$, where T_c is the folding temperature [18,20–23]. Yet, experimental data appear to yield a better fit to the scaling relation $n(t) \sim t^{1/\alpha}$ with $\alpha = (1+\nu)$ [19,24], where $\nu \simeq 0.588$ is the Flory exponent in three dimensions, which relates a polymer's radius of gyration in a good solvent to the number of monomers through $R_g \sim N^{\nu}$.

The mechanism for the observed anomalous dynamics has been investigated by exploiting the analogy to field-driven polymer translocation across a membrane (Y junction of the folding hairpin corresponding to the membrane pore) [17,25–27]. In fact, the exponent $(1+\nu)$ has been previously reported for the translocation time vs polymer length [28], under the assumption that the polymers on both sides of the pore are in quasiequilibrium at all times. Yet, as several studies pointed out [17,19], hairpin folding is an out-of-equilibrium phenomenon; therefore, observation of identical exponents in the two processes is conceivably coincidental. It was recently argued that the anomalous scaling of the hairpin folding time follows from Langevin dynamics under constant force, with a friction term associated with the relatively stretched portions of the unfolded arms [19]. While the time-limiting process in the folding dynamics is still unclear (further discussed below), the numerical value of α is also subject to continuing debate since out-of-equilibrium translocation processes and Monte Carlo simulations of hairpin folding on lattice models also show a regime with $\alpha = (1+2\nu)/(1+\nu)$ [17,27,29].

A marked difference between hairpin folding and polymer translocation phenomena is the rotational aspect of the dynamics, in the former case, induced by the natural twist of the DNA/RNA duplex. Despite past and recently renewed interest in statistical and dynamical properties of (un)winding polymers [30,31], existing studies on hairpin folding pay no attention to implications of duplex helicity. We here address the role of twist on zippering dynamics by comparing the folding rates of two computational models that are almost identical, except for the difference in angle and dihedral potentials, which induces an inherent twist in one model (helix) and but not in the other (ladder). By performing molecular dynamics (MD) simulations on chains more than an order of magnitude longer than the persistence length of the duplex, we demonstrate that twist is, in fact, an essential factor in

determining the folding-time scaling. As a bonus, the ladder model emerges, to our knowledge, as a unique example of hairpin folding that realizes the lower bound $\alpha = 2\nu$ imposed by energy conservation [26].

We use a coarse-grained one-bead-per-base model [see Figs. 1(a) and 1(b)] where a single DNA strand is held together by harmonic bonds with an equilibrium length b_0 and rotationally free joints at bead positions. Hairpin duplex is modeled to be composed of complementary bases occupying symmetric positions relative to the center. Base-pairing (interstrand) interaction is a segmented potential that has a minimum value of ϵ_{bp} at pair distance d_0, vanishes beyond a maximum bond distance $d_{\max}$, and is specific (each base is allowed to bond with its complement only). Pairing also induces interstrand angle and dihedral potentials that yield a DNA-like structure in the helical model in Fig. 1(a) and a zero-twist structure in the "ladder" model in Fig. 1(b). The local and nonlocal (excluded volume) potentials are used to enforce self-avoidance except for the "ghost" chain simulations. Associated potential functions and parameters are given in the Supplemental Material [32]. Hydrodynamic and electrostatic interactions are not included; hence Rouse dynamics is applicable, in line with the literature on the validity of Rouse versus Zimm regimes in the context of DNA dynamics [25,29,33–35].

MD trajectories are obtained by means of Langevin dynamics in an *NVT* ensemble, implemented in C ++ for speed. Prior to folding simulations, the critical temperature T_c was obtained separately for ladder and helix models by setting $\lambda(T_c) = 0.5$, where λ is the mean pair fraction. Contrary to one's "mechanical" intuition, the helical structure folds somewhat easier than the ladder ($T_c^l/T_c^h \simeq 1.08$) as a result of the smaller entropy of the helical duplex. The mismatch in duplex entropies is due to the difference in persistence lengths (34.3 bps and 23.3 bps for the helical and ladder models, respectively), a consequence of twist-bend coupling [36]. Note that, even the helical model is not a faithful representation of the actual DNA structure, but it captures the essential physical ingredients for the subject of this study and is simple enough to access long chains. Temperature in all of our MD simulations was chosen to be 0.16 in units of the pair bonding energy ϵ_{bp}, corresponding to $0.91T_c$ for the helix and $0.85T_c$ for the ladder model. Time was measured in dimensionless units $10^4 \times b_0\sqrt{m/\epsilon_{bp}}$ in all figures, where b_0 is the equilibrium bond length and m is the monomer mass.

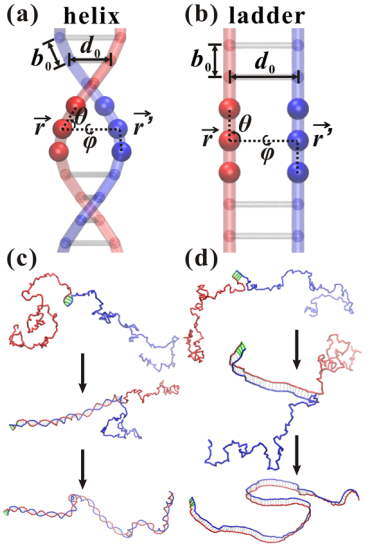

FIG. 1. Schematic view of the coarse-grained helix and ladder models (a), (b) and their folding processes (c), (d). θ and φ denote the interchain angle and dihedral potentials, respectively. b_0 and d_0 are the equilibrium bond lengths. $\vec{r}$ and $\vec{r'}$ are the radius vectors for the complementary bases.

In order to investigate folding dynamics [see Figs. 1(c) and 1(d) for snapshots] we followed the procedure outlined in Ref. [19]. In particular, the nucleation stage was bypassed by starting the MD simulations from an unfolded chain, which is equilibrated *a priori* at $T > T_c$ and has its first four base pairs (at the center of the polymer) permanently bound. The end effects were removed by defining the folding time τ as $n(\tau)/N = 0.75$. In order to elucidate the scaling behavior $\tau \sim N^{\alpha}$, we covered a wide range of hairpin lengths in the interval $24 \leq N \leq 644$. While lattice models can probe even larger systems sizes, they do not faithfully represent the helical duplex structure and associated rotational dynamics that we underline below.

Our central result is given in Fig. 2 where we plot both $n(t)$ vs t and the mean folding time τ vs N (all averaged over $\sim 10^3$ independent runs), separately for the helical and the ladder models. Data points spanning more than a decade in N are consistent with $\tau \sim N^{1+\nu}$ for the helical model. This exponent was reported earlier in Ref. [19] where significantly shorter chains were investigated by means of a three-beads-per-nucleotide hairpin model [37]. The ladder model obeys a visibly different scaling law that we postulate to be $\tau \sim N^{2\nu}$, also shown in Fig. 2 for comparison. The "whiskers" noticeable above the curves in Fig. 2 (also Fig. 3) reveal the speeding up during final stages of folding, where now shorter unbound segments facilitate faster relaxation.

The difference between helix and ladder folding times stands in contrast with the heuristic understanding for $\alpha = (1 + \nu)$ developed in Ref. [19], where the time evolution of the duplex length $n(t)$ was proposed to obey the Langevin equation $\gamma(n)\dot{n} = f$, with f a constant binding force satisfying $fb_0/k_BT \gtrsim 1$ and $\gamma(n) \sim n^{\alpha-1}$ representing the friction (up to logarithmic corrections [38]) on the stretched "stems" of the unfolded segments that are being pulled towards the Y junction. In this scenario, the folding speed is determined solely by the unpaired, single-strand portion of the hairpin. Accordingly, α is expected to remain

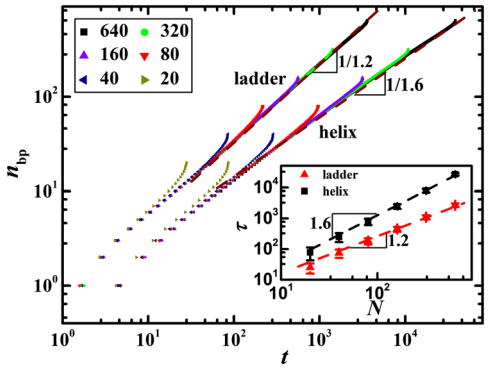

FIG. 2. Number of the formed base pairs as a function of time for the helix and ladder structures in three dimensions. The slopes are 0.63 ± 0.01 (helix) and 0.81 ± 0.01 (ladder), measured from the longest chain ($N = 640$) by using 17 equally spaced samples on the logarithmic scale. Wine colored dashed lines are the postulated slopes $1/1.6 = 0.625$ and $1/1.2 \simeq 0.83$ for the helix and the ladder, respectively. Inset shows the total folding time vs the length N of the hairpin.

unchanged after the helical duplex is replaced by a non-helical ladder geometry. The contrast between this expectation and our numerical results in Fig. 2 is striking, especially considering that a stretched "stem" can be observed in both cases (not shown). The new exponent $\alpha \simeq 1.2$ we find for "ladder folding" is also visibly different from $\alpha = (1 + 2\nu)/(1 + \nu) \simeq 1.37$ found in another out-of-equilibrium stress propagation model for DNA translocation through a pore [26]. A finite-size scaling analysis of the data for Fig. 2 and the subleading corrections it unveils are discussed in the Supplemental Material [32].

We propose that $\alpha = 2\nu$ for the ladder model and put it to test in two alternative settings where the Flory exponent is modified by (1) changing the dimension and (2) removing self-avoidance. As for (1), the obvious choice is to confine the ladder-like hairpin to two dimensions, since its equations of motion can trivially be constrained to a plane (in contrast, the finite thickness of the helical duplex makes confinement of the single-strand portions technically difficult). This is a convenient test ground where $(1 + \nu) = 1.75$ and $2\nu = 1.5$ are also easy to distinguish. The analysis of our MD simulations reported in Fig. 3 are in excellent agreement with $\alpha = 2\nu$.

Scenario (2) was implemented by removing the hard-core repulsion term in the model, hence producing a "ghost" polymer with $\nu = 0.5$. Interestingly, corresponding folding dynamics (shown in Fig. 4) is now ballistic *both for the ladder-like and the helical* hairpin models, where the duplex length $n(t)$ is proportional to the elapsed time with model-dependent growth rates. While ladder folding still conforms with $\alpha = 2\nu$, hence providing further support for our hypothesis, observation of the same exponent in the helical case disagrees with the above picture. Next, we address this issue.

An obvious difference between the dynamics of the two models is the rotational aspect of the folding process in the helical model. Since the single-stranded portion is much harder to rotate (except for the very last stages of folding), the helical duplex has to rotate around the centerline as it folds. Ladder model is not subject to such a constraint. We therefore check if the "ghost" helical model above is any different in this respect. The inset of Fig. 4 shows the rotation angle of the duplex around its centerline as a function of time for the original and the "ghost" helical models. In fact, in absence of self-avoidance, the helical duplex folds practically without any rotation; i.e., the "ghost" helical model is rotationally more similar to the ladder model. Hence, we conclude that the slow ($\alpha = 1 + \nu$) and fast ($\alpha = 2\nu$) folding behaviors observed above are linked to duplex rotation.

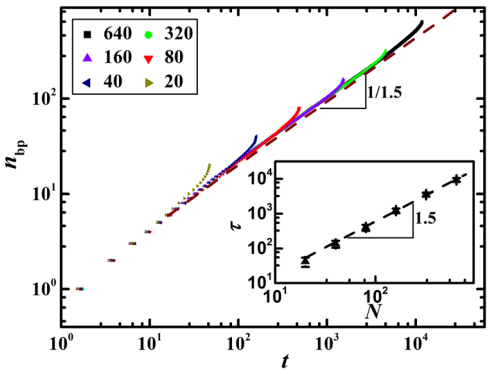

FIG. 3. Number of the formed base pairs as a function of time for the ladder model in two dimensions. The slope is 0.68 ± 0.01, measured from the longest chain ($N = 640$) using 17 equally spaced samples on the logarithmic scale. Wine colored dashed line is the postulated slope of $1/1.5 \simeq 0.67$. Inset shows the total folding time vs the length N of the hairpin.

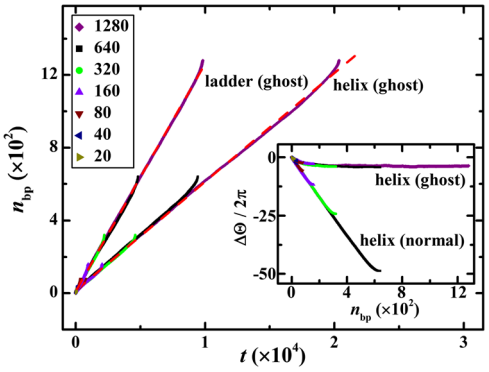

FIG. 4. Number of the formed base pairs as a function of time for the helix and ladder structures using a "ghost" model in three dimensions. The red color dashed lines correspond to constant folding rates. The inset shows the comparison of the rotation angle ($\Delta\Theta$) of duplex (about the centerline) in the "ghost" and normal (self-avoiding) helical models.

It is worthwhile to point out that $\alpha = 2\nu$ is a lower bound set by the folding energetics. To show this, we adopt an argument from polymer translocation studies [26] to the present context: given that $\tau(s) \sim s^{\alpha}$, the mean velocity of the bases with index s (between the beginning of the folding process and the instant they pair up) is

$$v(s) \sim \frac{|\vec{r}(s) - \vec{r'}(s)|}{s^{\alpha}} \sim s^{\nu-\alpha}, \tag{1}$$

where $\vec{r}$ and $\vec{r'}$ are the radius vectors for the complementary bases. The energy lost to friction can then be expressed as

$$E_f \sim \int_0^N \eta v(s) s^{\nu} ds \sim \eta N^{2\nu-\alpha+1}, \tag{2}$$

where η is the friction coefficient and s^{ν} is the distance travelled by the s^{th} pair until they bind, so that the integrand is the work done against friction. The sum total of this work is supplied by N pair-binding events during folding; therefore we expect

$$N^{2\nu-\alpha+1} \leq N \quad \text{or} \quad \alpha \geq 2\nu. \tag{3}$$

Our observations suggest that hairpin folding is a complex nonequilibrium phenomenon involving both rotational and translational relaxation processes. In absence of helicity, folding progresses at a speed limited only by the constraint on the rate of energy transfer between bonds forming at the Y junction and the viscous environment. On the other hand, the folding rate of an DNA-like hairpin structure is significantly slower due to the required rotational relaxation of the duplex (as implied by Fig. 4 inset).

A polymer translocating through a pore also displays a fast and a slow regime, depending on the applied force. Translocation time (vs length) exponents $\alpha = (1 + 2\nu)/(1 + \nu)$ (attributed to stress propagation dynamics) for fast translocation and $\alpha = (1 + \nu)$ for slow translocation [29] are analogous to the hairpin folding scenario here. In Fig. 5 we show that, in the slow-folding regime, the helical duplex remains not far from equilibrium during the entire folding process, while the fast-folding ladder structure maintains a duplex that is very compact (more so than the trans portion in fast translocation simulations [29]).

Motivated by these observations, below we propose a scaling argument for $\alpha = (1 + \nu)$ in helical hairpin folding. Given that the relaxation of the duplex is the rate-limiting step, we assume that the duplex with length $n(t)$ has size $\sim n(t)^{\nu}$ at all times (see Fig. 6). A Langevin equation for the folding process can then be written as

$$\eta(n) \frac{dx}{dt} \sim f, \tag{4}$$

where $\eta(n) \sim n$ is the friction coefficient; dx is the displacement of the duplex due to an added pair as depicted in Fig. 6,

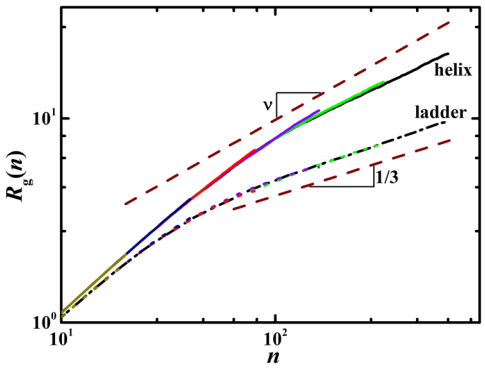

FIG. 5. Radius of gyration of the duplex vs the duplex length during the folding processes of the helix and ladder in three dimensions. Wine colored dashed lines indicate the upper bound (ν) and lower bound ($1/3$) for the slopes corresponding to equilibrated and fully compact structures, respectively.

$$dx = (n + dn)^{\nu} - n^{\nu} \sim \nu n^{\nu-1} dn, \tag{5}$$

and f is the constant force at the Y junction due to the binding potential. Then Eq. (4) becomes

$$n n^{\nu-1} dn \sim dt. \tag{6}$$

Integrating Eq. (6) yields the folding time, $\tau \sim N^{1+\nu}$. Note that, unlike the polymer translocation problem, the mechanical properties of the polymer on the two sides of the Y junction are quite different. This introduces different relaxation timescales, allowing one side (duplex) to maintain a quasiequilibrium state throughout most of the folding, while the other (unpaired) segments are out of equilibrium. In the ladder scenario, the relaxation of the duplex is not a prerequisite for pairing, hence folding progresses with both sides visibly out of equilibrium. In this case, we observed that the folding rate is limited only by the allowed rate of energy flow to the fluid, given the initial equilibrium configuration of the hairpin. In conclusion, our findings provide a novel perspective on hairpin folding dynamics by unveiling the significant, and so far ignored, contribution of the rotational motion of the duplex in the process.

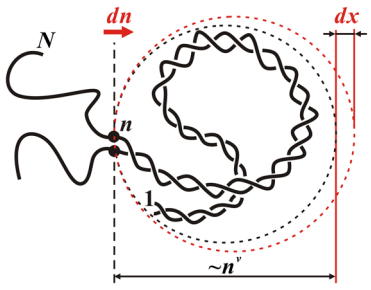

FIG. 6. A cartoon depiction of the hairpin folding dynamics and the notation used in Eqs. (4)–(5). Dark circles label the last bound pair.

We are thankful to E. Carlon for his helpful comments on our preliminary results. We are also in debt to M. Öztürk and M. Sayar for their contributions during model development, as well as two anonymous reviewers for a significant improvement of the manuscript. This work is supported by the Scientific and Technological Research Council of Turkey (TÜBITAK) through Grant No. MFAG-114F348.

*akabakcioglu@ku.edu.tr

[1] X. Zhuang, L. E. Bartley, H. P. Babcock, R. Russell, T. Ha, D. Herschlag, and S. Chu, Science **288**, 2048 (2000).
[2] H. F. Noller, Science **309**, 1508 (2005).
[3] T. Pan and T. Sosnick, Annu. Rev. Biophys. Biomol. Struct. **35**, 161 (2006).
[4] S. V. Solomatin, M. Greenfeld, S. Chu, and D. Herschlag, Nature (London) **463**, 681 (2010).
[5] G. Peters, P. Coussement, J. Maertens, J. Lammertyn, and M. De Mey, Biotechnology advances **33**, 1829 (2015).
[6] S. Teotia, D. Singh, X. Tang, and G. Tang, Trends Biotechnol. **34**, 106 (2016).
[7] K. Kim, S. M. Ryu, S. T. Kim, G. Baek, D. Kim, K. Lim, E. Chung, S. Kim, and J. S. Kim, Nat. Biotechnol. **35**, 435 (2017).
[8] S. F. Dowdy, Nat. Biotechnol. **35**, 222 (2017).
[9] S. Rouskin, M. Zubradt, S. Washietl, M. Kellis, and J. S. Weissman, Nature (London) **505**, 701 (2014).
[10] A. Solem and A. Laederach, Nat. Chem. Biol. **11**, 906 (2015).
[11] D. B. Ritchie and M. T. Woodside, Curr. Opin. Struct. Biol. **34**, 43 (2015).
[12] W. Zhang and S.-J. Chen, Proc. Natl. Acad. Sci. U.S.A. **99**, 1931 (2002).
[13] J. Jung and A. Van Orden, J. Am. Chem. Soc. **128**, 1240 (2006).
[14] H. Ma, C. Wan, A. Wu, and A. H. Zewail, Proc. Natl. Acad. Sci. U.S.A. **104**, 712 (2007).
[15] J. Jung, R. Ihly, E. Scott, M. Yu, and A. Van Orden, J. Phys. Chem. B **112**, 127 (2008).
[16] A. Ansari, S. V. Kuznetsov, and Y. Shen, Proc. Natl. Acad. Sci. U.S.A. **98**, 7771 (2001).
[17] A. Ferrantini and E. Carlon, J. Stat. Mech. 2011, P02020.
[18] T. Sakaue, J.-C. Walter, E. Carlon, and C. Vanderzande, Soft Matter **13**, 3174 (2017).
[19] R. Frederickx, T. intVeld, and E. Carlon, Phys. Rev. Lett. **112**, 198102 (2014).
[20] S. Cocco, J. F. Marko, and R. Monasson, Eur. Phys. J. E **10**, 153 (2003).
[21] C. Richard and A. J. Guttmann, J. Stat. Phys. **115**, 925 (2004).
[22] D. Poland and H. A. Scheraga, J. Chem. Phys. **45**, 1456 (1966).
[23] M. E. Fisher, J. Chem. Phys. **45**, 1469 (1966).
[24] K. Neupane, D. B. Ritchie, H. Yu, D. A. N. Foster, F. Wang, and M. T. Woodside, Phys. Rev. Lett. **109**, 068102 (2012).
[25] M. Manghi and N. Destainville, Phys. Rep. **631**, 1 (2016).
[26] H. Vocks, D. Panja, G. T. Barkema, and R. C. Ball, J. Phys. Condens. Matter **20**, 095224 (2008).
[27] V. V. Palyulin, T. Ala-Nissila, and R. Metzler, Soft Matter **10**, 9016 (2014).
[28] Y. Kantor and M. Kardar, Phys. Rev. E **69**, 021806 (2004).
[29] K. Luo, T. Ala-Nissila, S. C. Ying, and R. Metzler, Europhys. Lett. **88**, 68006 (2009).
[30] M. Baiesi, G. T. Barkema, E. Carlon, and D. Panja, J. Chem. Phys. **133**, 154907 (2010).
[31] J. C. Walter, M. Baiesi, E. Carlon, and H. Schiessel, Macromolecules **47**, 4840 (2014).
[32] See Supplementary Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.121.138101 for force field and the corresponding parameters for the coarse-grained DNA model and the finite-size scaling analysis of the folding dynamics.
[33] T. Ikonen, A. Bhattacharya, T. Ala-Nissila, and W. Sung, Europhys. Lett. **103**, 38001 (2013).
[34] A. Izmitli, D. C. Schwartz, M. D. Graham, and J. J. de Pablo, J. Chem. Phys. **128**, 085102 (2008).
[35] J. Smiatek, Ph. D. thesis, Bielefeld Univ., Germany (2009).
[36] J. F. Marko and E. D. Siggia, Macromolecules **27**, 981 (1994).
[37] E. Sambriski, D. Schwartz, and J. de Pablo, Biophys. J. **96**, 1675 (2009).
[38] T. Sakaue, Phys. Rev. E **76**, 021803 (2007).