Tightness of random knotting

Vsevolod Katritch,[1,*] Wilma K. Olson,[1] Alexander Vologodskii,[2] Jacques Dubochet,[3] and Andrzej Stasiak[3,†]
[1]*Department of Chemistry, Rutgers, The State University of New Jersey, New Brunswick, New Jersey 08903*
[2]*Department of Chemistry, New York University, New York, New York 10003*
[3]*Laboratoire d'Analyse Ultrastructurale, Bâtiment de Biologie, Université de Lausanne, CH-1015 Lausanne-Dorigny, Switzerland*
(Received 19 February 1999; revised manuscript received 15 February 2000)

Long polymers in solution frequently adopt knotted configurations. To understand the physical properties of knotted polymers, it is important to find out whether the knots formed at thermodynamic equilibrium are spread over the whole polymer chain or rather are localized as tight knots. We present here a method to analyze the knottedness of short linear portions of simulated random chains. Using this method, we observe that knot-determining domains are usually very tight, so that, for example, the preferred size of the trefoil-determining portions of knotted polymer chains corresponds to just seven freely jointed segments.

PACS number(s): 61.41.+e, 05.40.Fb, 83.10.Nn

A topological perspective dominates current thinking about knotted polymers [1,2]. Knotted configurations obtained by numerical approaches based on random-walk simulations [3–7] and experimentally obtained knots [4,8] are typically characterized in terms of knot types and/or global physical properties. The comparable average sizes of very long knotted and unknotted circular trajectories simulated on a cubic lattice [9,10] as well as entropy considerations [11] suggest the presence of tight knots in long, randomly distorted polymers. However, other numerical studies [12] were not supporting the preference for tight knotting.

To our knowledge, the question of whether entanglements tend to be ''spread'' over the whole molecule or localized within a short region has never been directly analyzed; see Fig. 1 for a schematic illustration of these two extremes. Also, the typical extent of tightness of random knots has not been investigated before. We decided, therefore, to simulate trajectories adopted by knotted polymeric chains freely suspended in a solvent under conditions where the segments neither attract nor repel one another and then to measure the size of the knot-determining domains for different types of knots made up of different length polymers. We first collected ensembles of independent knotted configurations from sets of 500 000 or more closed random walks of a given length using the simulation method described in Refs. [13, 14], distinguishing different knot types on the basis of their Alexander polynomials [15–17]. The chain length is measured in terms of the number of freely jointed (Kuhn) segments, each segment corresponding to the real length over which the correlation between the starting and ending directions of the polymer is completely lost, e.g., ~300 base pairs in double-stranded DNA [18,19]. It was conjectured [20,21] and rigorously proven for lattice and continuum models of polymer behavior [22,23] that the probability of knotting tends to unity for very long walks. Numerous simulation studies and cyclization of long DNA molecules [3–5,7,8] addressed previously the probability of knotting as a function of the size of random walks. Figure 2 shows that the probability of remaining unknotted drops exponentially to ~1% as chain length grows from 50 to 1500 freely jointed segments, while the formation of simple prime knots reaches a maximum in chains of ~260 segments. For still longer chains the competition between the increasing number of realizable knots (particularly composite knots) brings about a gradual decrease in the probability of forming a given prime knot; see also Ref. [5]. The designation of the knot types in Fig. 2 follows the standard notation in which the main number indicates the minimal number of crossings that a given knot can have in a planar projection and the subscript indicates the tabular position of this knot among all knots with the same minimal number of crossing [16,17]. The composite knot notation lists the two or more independent factor knots tied on the same string. Regularized configurations of the analyzed knots can be seen in Refs. [16,17,24,25].

The frequency of knotting versus chain length does not directly tell us whether polymer knots are loose, i.e., distributed over the whole length of the modeled walk [Fig. 1(a)], or tight, i.e., localized within a short portion of the molecule [Fig. 1(b)]. To identify and measure the size of knotted domains, we systematically scanned the generated configurations by providing an external closure to a ''moving window'' containing a given number of segments. Conceptually similar types of closures were proposed earlier to detect knotting in linear random chains [26]. We looked for the minimal number of segments that, upon closure with an external planar loop, formed the same type of knot as the entire circular chain. Knot-type recognition was based on calculations of the Alexander polynomial for complete circular walks and for selected length linear subchains that were closed with external planar loops. A large planar loop forming almost a complete circle was attached to the starting and ending vertices of the analyzed subdomain. The actual inclination of the planar loop with respect of the enclosed subchain was random since local orientation does not affect the topology of the construct. The diameter of the closing loop was chosen to be much larger than the dimensions of the analyzed subdomain in order to minimize the risk of piercing through the enclosed portion of the chain and thereby chang-

*Present address: The Scripps Research Institute, Mail MB-37, La Jolla, CA 92037.
†Author to whom correspondence should be addressed. Electronic address: Andrzej.Stasiak@lau.unil.ch

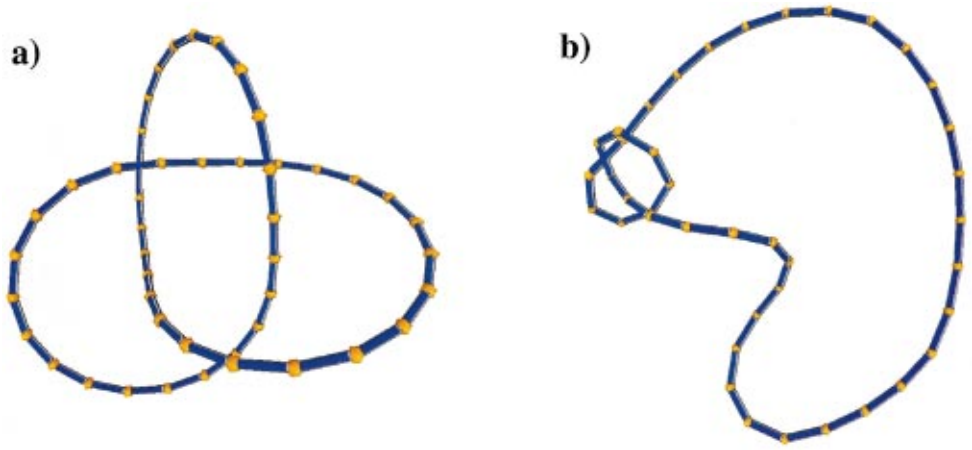

FIG. 1. (Color) Schematic illustration of (a) loose and (b) tight trefoil (3_1) knotting. Notice that real and simulated polymer chains at equilibrium do not have such smooth, regular trajectories (see Fig. 3 for one such simulated trajectory).

ing the knot type of the total construct. Accidental piercing is a problem during the closure of long subchains but is very rare for small enclosed subchains. Figure 3 illustrates detection of the knotted domain in a randomly generated, 100-segment knotted chain. The detected trefoil-determining domain, which is composed of seven segments, is evident as is the unknotted character of the remainder of the chain. Only part of the 200-segment long virtual arc attached to the boundaries of the knot-determining domain is shown in the figure.

While knotting events induced by closure with an external loop are rare, it is important to see whether these false knot-

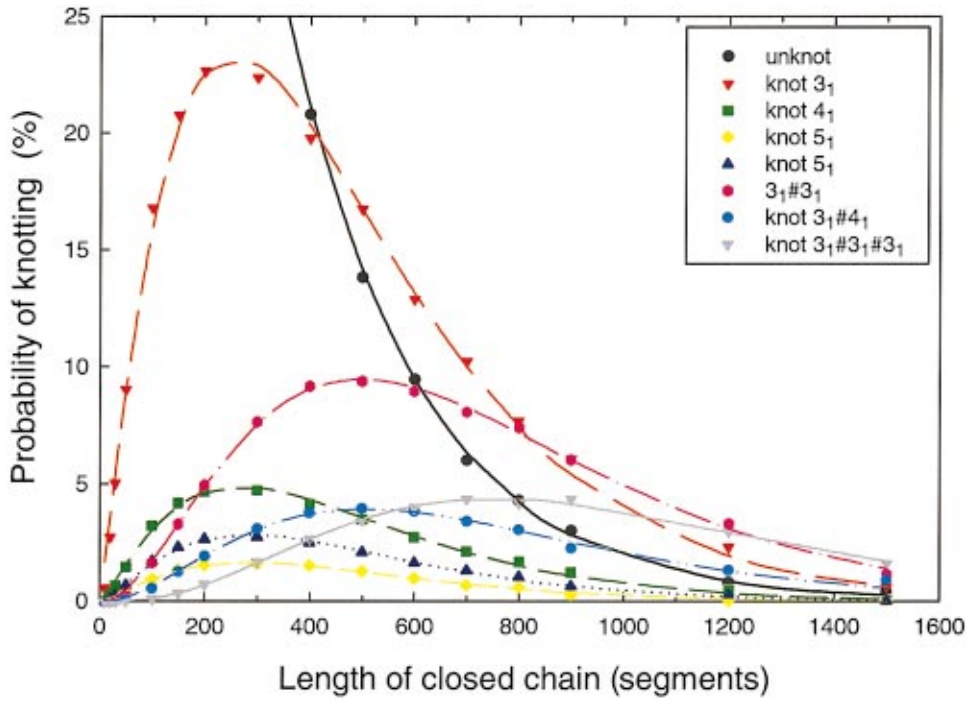

FIG. 2. (Color) Probability of forming simple prime and composite knots as a function of the number N of freely jointed chain segments in simulated closed polymers. The smooth curves are fits of the computed probabilities to an exponential, $p = aN^b \exp(-N/c)$. The composite knots ($3_1\#3_1$, $3_1\#4_1$, and $3_1\#3_1\#3_1$) are described in terms of the prime knots [16,17] that comprise them. Thus, the notation $3_1\#4_1$ denotes a knot built from a trefoil (3_1) and an achiral four-noded (4_1) knot [24]. Random closed chains were simulated using the algorithm described in Refs. [13,14]. Because the Alexander polynomial does not distinguish between mirror images of chiral knots such as 3_1, 5_1, and 5_2, the right- and left-handed versions of these knots are grouped together. For similar topological reasons, the sample of $3_1\#3_1$ composite knots includes a very small proportion of 8_{20} knots with the same Alexander polynomial [9]. While other more complex knots can have the same Alexander polynomial as the simple 3_1, 4_1, 5_1, and 5_2 knots analyzed by us, the occurrence of such knots is so rare that they will not affect the statistics in an appreciable way and are most likely absent from our samples. Qualitatively similar findings were reported previously by Deguchi and Tsurusaki [5]. The ''step number'' associated with the ~10 000 random chains generated in that work, however, does not correspond to the number of freely jointed segments reported here. The present results are also based on a better sampling with more than 500 000 independent closed configurations analyzed at each chain length.

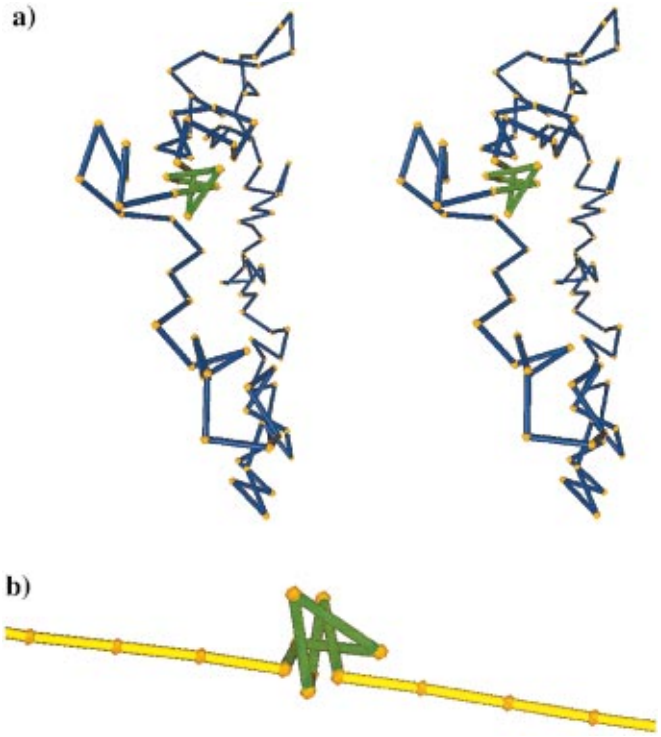

FIG. 3. (Color) Example of a tight trefoil-determining domain identified in a random configuration of a trefoil forming polymer: (a) stereo representation of a seven-segment trefoil, colored in green, in the context of a 100-segment chain; (b) a virtual quasicircular planar arc is attached to the ends of the seven segments shown above. Alexander polynomial calculation applied to this construct (seven segments of the chain and the virtual closure) detects a 3_1 knot. Careful visual analysis of the figure also confirms the presence of a trefoil knot.

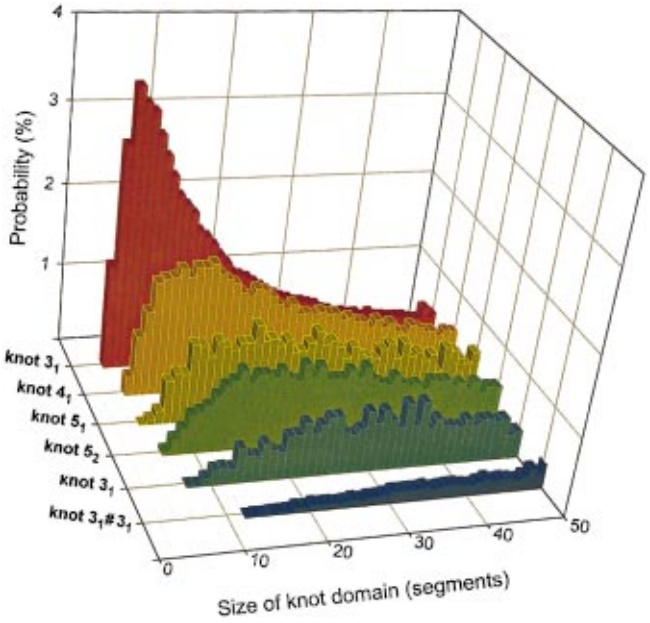

FIG. 4. (Color) Sizes of knot-determining domains found in computer simulated freely jointed chains composed of 500 segments. The data were obtained by scanning thousands of $(3_1,4_1,5_1,5_2,6_1)$ prime and $(3_1\#3_1)$ composite knots collected from a set of 900 000 random circular walks of 500 steps (see Table I for the corresponding numbers of analyzed knots). The length of the knot-determining domain in each of the individual configurations is defined as the smallest number of segments which, upon virtual closure (Fig. 3), produces the knot type of the whole knotted chain. The smoother histograms of some knot types $(3_1,4_1,3_1\#3_1)$ reflect the greater numbers of those knots in our sample; see Table I. Notice that for each knot type, 100% corresponds to the total number of analyzed trajectories of a given knot type.

ting events significantly affect our results. To this aim we performed a control of our knot detection method by carrying out a ''knotted domain search'' on unknotted closed chains. The proportion of incorrectly detected trefoil domains on 500-segment unknotted rings that resulted from additional crossings by an external closing loop was ~3% in a sample of ~100 000 chains. The distribution of minimal sizes among these false knotted domains, however, was strongly shifted toward much larger values than in the corresponding set of knotted domains detected in real knotted chains. False knotting occurs at larger chain lengths, since only for larger subdomains is there a significant chance that the closing procedure will introduce topological entanglements. Real knots are detected in smaller subdomains, because they do not need additional entanglements. Thus, while a small proportion of the data for trefoil knots in Fig. 4 may result from false knotting events, the position of the maximum is unlikely to be affected by false knotting. False detection of 4_1 knots is of the order of 1%, while that of more complex knots is even lower.

The histograms in Fig. 4 show the distributions of the minimal number of segments in knot-determining domains within 500-segment circular random walks forming 3_1, 4_1, 5_1, 5_2, and 6_1 prime knots and in $3_1\#3_1$ composite knots made up of two trefoils. The most frequent minimal size for the trefoil-determining domains is 6–8 segments, and the observed maximum of this distribution is well pronounced. As could be expected, the domains determining 4_1 knots span more segments with the most frequent minimal realizations of such domains requiring ~15 segments, and the maximum of the distribution being less pronounced than that of the trefoil domains. For 5_1 and 5_2 knots we see further broadening of the distributions with the most frequent minimal size of the domains determining these knots at ~20 segments, while the most frequent minimal size of 6_1 knots reaches more than 30 segments. As can be seen in Fig. 4, the most frequent minimal size of prime knots grows progressively with the complexity of these knots. We showed earlier that the length to diameter ratio of ideal geometric configurations of knots is a good measure of the knots' complexity [24]. We checked, therefore, how the length/diameter ratio of ideal knots relates to the most frequent minimal size of random knots of the corresponding size. For the prime knots analyzed here (3_1, 4_1, 5_1, 5_2, and 6_1), this relation is close to linear. However, $3_1\#3_1$ composite knots display qualitatively different behavior consistent with independent localization of both factor knots [10,27]. For $3_1\#3_1$ composite knots, a minimal size of more than 50 segments is needed to enclose both factor knots, although this knot in its ideal geometric configuration has a very similar length/diameter ratio to the 6_1 knot [24,25].

In addition to measuring the distribution of minimal sizes of knot-determining domains, we have checked the proportion of confined and redistributed configurations of different

TABLE I. Frequency of confinement of knots-determining domains into $\frac{1}{10}$ size of 500-segment-long closed random walks (in a set of 900 000 walks).

Knot type	Number of analyzed trajectories	Confined knots (%)
3_1	151278	55.6
4_1	32420	44.6
5_1	11199	33.8
5_2	19273	32.5
6_1	5938	21.4
6_2	6420	20.4
6_3	3722	17.5
$3_1\#3_1$	85895	5.1
$3_1\#4_1$	35809	3.4

knots. As a somewhat arbitrary criterion, we chose to measure the proportion of knots that are localized in $\frac{1}{10}$ of the contour length of a knotted polymer. As shown in Table I, more than 55% of the trefoil-determining domains of 500-segment trefoil-forming chains are confined within this limit, i.e., 50 segments or fewer. For more complicated knots this proportion becomes smaller but is still as large as 20% for knots with six crossings. As expected, the chance of the two trefoil-determining domains of 500-segment, composite $3_1\#3_1$ knots falling within a 50-segment window is very small (5%), a result confirming the independent localization of both factor knots [10,27].

It follows that relatively simple knots such as 3_1, 4_1, 5_1, and 5_2 prime knots tend to be confined rather than distributed over circular polymers of several hundred freely jointed segments. To understand the physical basis of this confinement, it is useful to consider first the probability of simple circularization of a polymer as a function of chain length. The polymer should be long enough that the overall bending energy required to close the chain can be furnished by thermal fluctuations. On the other hand, the chain should not be so long that the chance that its two ends will meet gets very small. The average distance between the ends of a thermally agitated polymer increases with chain length, and both the relative concentration of one end of the polymer in the vicinity of the other and the probability of circularization decrease [19]. Thus polymer circularization occurs at an optimum chain length in which the two ends of the molecule are not too far apart but where the chain is sufficiently long to undergo thermally induced 360° bending. In the case of a freely jointed polymer, this chain length corresponds to approximately two statistical segments [28,29].

If we apply the above reasoning to the probability of formation of a trefoil-determining domain by a growing polymer, we conclude that the chain must be sufficiently long to allow the thermally induced bending required for the formation of the knotted domain. In addition, since the total curvature of a knot is much greater than that of an unknot [30], both the bending and the required length of the polymer should be significantly greater than the corresponding values needed for simple circularization. At the same time, we note that stochastically formed loops are most likely to be pierced by a growing chain end which is not far away. As soon as the growing end becomes distant from the loop, the chance that it will pierce this subdomain and form a knot decreases. Thus knotting events will tend to be confined to relatively short chain lengths.

Our observation of tight knotting may at first seem incompatible with the results in Fig. 2, which demonstrate that very long polymers (corresponding to ~250 freely jointed segments) are needed to achieve a substantial probability of trefoil formation. While polymer lengths corresponding to seven segments are the most frequent size of trefoil-determining domains, the actual probability that a given seven statistical segments will form a trefoil-determining domain is quite low. Therefore, a polymer must sample many independent configurations of seven freely jointed segments to achieve a substantial probability of trefoil formation. The observed tightness is a ''local'' characteristic of trefoil-forming polymers, whereas the lengths in Fig. 2 reflect the most likely ways of incorporating this topological constraint in a long, closed circular molecule.

Although we have concentrated in this work on the size of knot-determining domains in closed chains of a given knot type, our results can be extended to long linear chains which accumulate tightly entangled domains [26,31]. Here we have focused on the knotting of chains of zero thickness which reflect the ideal behavior of polymers under so-called θ conditions [32,33]. After this work was completed, we learned about a study which shows that knots also tend to localize within small domains of long self-avoiding random walks [34]. Self-avoiding random walks mimic the behavior of polymers subject to so-called excluded volume effects which act on polymers suspended in good solvents [32,33].

We thank Alexander Grosberg and John Marko for discussions. This work was supported by the U.S. Public Health Service (Grants Nos. GM34809 and GM54215), the Burroughs-Wellcome Fund through the Program in Mathematics and Molecular Biology based at Florida State University, the Swiss National Foundation (Grant No. 31-42158), and the Foundation Herbette of the University of Lausanne.

[1] J. H. White and N. R. Cozzarelli, Proc. Natl. Acad. Sci. USA **81**, 3322 (1984).
[2] D. W. Sumners, Mathematical Intelligencer **12**, 71 (1990).
[3] A. V. Vologodskii *et al.*, Zh. Eksp. Teor. Fiz. **66**, 2153 (1974) [Sov. Phys. JETP **39**, 1059 (1974)].
[4] V. V. Rybenkov, N. R. Cozzarelli, and A. V. Vologodskii, Proc. Natl. Acad. Sci. USA **90**, 5307 (1993).
[5] T. Deguchi and K. Tsurusaki, J. Knot Theory and Its Ramifications **3**, 321 (1994).
[6] M. C. Tesi *et al.*, Phys. Rev. E **49**, 868 (1993).
[7] K. Koniaris and M. Muthukumar, Phys. Rev. Lett. **66**, 2211 (1991).
[8] S. Y. Shaw and J. C. Wang, Science **260**, 533 (1993).
[9] E. J. Janse van Rensburg and S. G. Whittington, J. Phys. A **24**, 3935 (1991).
[10] E. Orlandini, M. C. Tesi, E. J. Janse van Rensburg, and S. G.

Whittington, J. Phys. A **31**, 5953 (1998).
[11] A. Y. Grosberg, A. Feigel, and Y. Rabin, Phys. Rev. E **54**, 6618 (1996).
[12] S. R. Quake, Phys. Rev. E **52**, 1176 (1995).
[13] K. V. Klenin *et al.*, J. Biomol. Struct. Dyn. **5**, 1173 (1988).
[14] A. V. Vologodskii and M. D. Frank-Kamenetskii, *Modeling DNA Supercoiling*, in *Methods in Enzymology*, edited by D. M. J. Lilley and J. Dahlberg (Academic Press, San Diego, 1992), p. 467.
[15] J. W. Alexander, Trans. Am. Math. Soc. **30**, 275 (1928).
[16] D. Rolfsen, *Knots and Links* (Publish or Perish, Berkeley, CA, 1976).
[17] C. C. Adams, *The Knot Book* (Freeman, New York, 1994).
[18] S. D. Levene and D. M. Crothers, J. Mol. Biol. **189**, 73 (1986).
[19] A. V. Vologodskii, *Topology and Physics of Circular DNA*, *Physical Approaches to DNA* (CRC, Boca Raton, FL, 1992).
[20] H. L. Frisch and E. Wasserman, J. Am. Chem. Soc. **83**, 3789 (1961).
[21] M. Delbrück, Knotting Problems in Biology, in *Mathematical Problems in the Biological Sciences*, edited by R. E. Bellman (American Mathematical Society, Providence, RI, 1962), p. 55.
[22] D. W. Sumners and S. G. Whittington, J. Phys. A **21**, 1689 (1988).
[23] Y. Diao, N. Pippenger, and D. W. Sumners, J. Knot Theory Ramifications **3**, 419 (1994).
[24] V. Katritch *et al.*, Nature (London) **384**, 142 (1996).
[25] V. Katritch *et al.*, Nature (London) **388**, 148 (1997).
[26] E. J. Janse van Rensburg, D. W. Sumners, E. Wasserman, and S. G. Whittington, J. Phys. A **25**, 6557 (1992).
[27] Y.-J. Sheng, P.-Y. Lai, and H.-K. Tsao, Phys. Rev. E **58**, R1122 (1998).
[28] H. Yamakawa and W. H. Stockmayer, J. Chem. Phys. **57**, 2843 (1972).
[29] D. Shore and R. L. Baldwin, J. Mol. Biol. **170**, 957 (1983).
[30] J. W. Milnor, Ann. Math. **52**, 248 (1950).
[31] R. Matthews, Math. Today **33**, 82 (1997).
[32] P. J. Flory, *Principles of Polymer Chemistry* (Cornell University Press, Ithaca, NY, 1953).
[33] P.-G. de Gennes, *Scaling Concepts in Polymer Physics* (Cornell University Press, Ithaca, NY, 1979).
[34] E. Guiter and E. Orlandini, J. Phys. A **32**, 1359 (1999).