Limits of analogy between self-avoidance and topology-driven swelling of polymer loops

N. T. Moore and A. Y. Grosberg
Department of Physics, University of Minnesota, Minneapolis, Minnesota 55455, USA
(Received 30 June 2005; published 21 December 2005)

The work addresses the analogy between trivial knotting and excluded volume in looped polymer chains of moderate length, where the effects of knotting are small. A simple expression for the swelling seen in trivially knotted loops is described and shown to agree with simulation data. Contrast between this expression and the well-known expression for excluded volume polymers leads to a graphical mapping of excluded volume to trivial knots, which may be useful for understanding where the analogy between the two physical forms is valid. The work also includes description of a new method for the computational generation of polymer loops via conditional probability. Although computationally intensive, this method generates loops without statistical bias, and thus is preferable to other loop generation routines in the region of interest.

DOI: 10.1103/PhysRevE.72.061803 PACS number(s): 61.25.Hq

I. INTRODUCTION: FORMULATION OF THE PROBLEM

The last few years have seen significant work addressing the effects of knotting on looped polymer chains. Of interest to mathematicians and physicists for a good part of nineteenth and most of the twentieth centuries, knots were seen by W. Thomson as a way to understand the nature of atoms [1], and more recently as the basis for string theory. On the biological front, knots have been observed in [2,3] and tied into [4,5] strands of DNA. Additionally, topoisomerases—proteins which act to alter the topological state of DNA—are quite common and play a significant role in cellular processes.

The most obvious effect knotting has on a loop is in the size, commonly measured in terms of radius of gyration, R_g^2. For instance, the loop that is topologically equivalent to a circle, called a trivial or 0_1 knot in professional parlance, is on average *larger* than the loop of the same length with any other topology. In other words, a trivial loop is larger than the phantom loop, the latter representing a topology-blind average over all loops of a certain length: $\langle R_g^2\rangle_{triv} > \langle R_g^2\rangle_{phantom}$.

It is difficult to approach knots theoretically, because topological constraints are fundamentally nonlocal, even though they are caused by the same short-range repulsion forces between monomers as classical self-avoidance. Luckily, there is a wide range of parameters in which the effect of self-avoidance is only marginal, while knots are clearly present. For instance, for dsDNA, because of its low flexibility, one can safely ignore swelling due to the excluded volume for the length up to about several thousands Kuhn segments, while knots can of course exist in much shorter molecules. A very interesting question arises when one thinks about knots in polymers under θ conditions. In this case, the swelling due to the excluded volume is compensated by the attractive forces, while the latter have no influence on the topological constraints.

The important conclusion is that the topology-driven swelling is operational even for very thin polymers, in the limit when volume exclusion has no effect on polymer coil size. In this case, the phantom loop's size (which is, once again, average over all topologies) scales as $N^{1/2}$, while the trivial loop is larger not merely because of a larger prefactor, but because of a larger scaling exponent, its size scales as N^{ν}, where $\nu > \frac{1}{2}$. The conjecture, formulated a long time ago [6], supported by further scaling arguments [7,8], and consistent with recent simulation data [9–11], specifies that the scaling exponent ν describing topology-driven swelling of a trivial loop is exactly the same as the Flory exponent [12], which describes swelling driven by the self-avoidance (or excluded volume): $\nu \approx 0.589 \approx \frac{3}{5}$.

Equality of scaling exponents for the two cases reflects the similarity of fractal properties for these systems at very large $N \gg 1$, because topological constraints result in self-avoidance of blobs on all length scales above a certain threshold [8]. As we understand much about self-avoidance [13], and next to nothing about knots, we would like to exploit the analogy to see if it yields any insights into knots. Specifically, it is tempting to look at the dependence of the unknotted loop size, $\langle R_g^2\rangle_{triv}$, on the number of segments N, not only in the asymptotic scaling regime of very large N, but also the corrections to scaling at not-so-large N. This is particularly important from a practical standpoint, because the asymptotic scaling limit is barely accessible computationally, and what one really computes is the value of $\langle R_g^2\rangle_{triv}$ at rather moderate N. Systematic comparison of N dependencies of $\langle R_g^2\rangle$ for (trivial) knots and self-avoiding polymers over the wide range of N is the goal of this paper.

We show that although large N scaling appears to be identical for trivial knots and excluded volume polymers, their respective approach to the asymptotic regime is different. This points obviously to the limited character of the analogy between the two mechanisms of swelling, due to volume exclusion and due to topological constraints.

The plan of the paper is as follows. We start from a brief summary of the main results for self-avoiding polymers. Although these results are widely known, we restate them in the form most suitable for our purposes. Next, we present some heuristic analytical arguments to shed light on why trivial knots may behave differently than their excluded volume counterparts. With this insight in mind, we present our detailed computational data on the N dependence of $\langle R_g^2\rangle_{triv}$

over the wide range of N. To obtain data with the necessary degree of accuracy, it is necessary to make sure that our method of generating loops is ergodic and unbiased. Although this aspect is of decisive importance, it is purely technical, and thus it is relegated to the Appendix. Up to about Sec. II C, we mostly review the known results, starting from Sec. II D, we present our findings.

II. PRELIMINARY CONSIDERATIONS

A. Swelling driven by self-avoidance: An overview

To make our work self-contained, we now offer a brief review of the results for the scaling of excluded volume polymers (see further details in Refs. [13–15]). We should emphasize from the beginning that the main properties of the excluded volume polymer are valid also for loops [16]. The simplest model for excluded volume is a system in which N beads, each of volume b, are placed along a loop with mean separation ℓ. All other forms of excluded volume, e.g., freely jointed stiff rods, wormlike filaments, etc., can be mapped to this simple rod-bead model (see, e.g., Ref. [14]). There are two scaling regimes, with crossover at the length

$$N^* \sim (\ell^3/b)^2. \tag{1}$$

In terms of N^*, the mean squared gyration radius $\langle R_g^2 \rangle$ can be written as $\langle R_g^2 \rangle = \ell^2 N \rho(z)$, where the swelling factor ρ depends on the single variable $z = \sqrt{N/N^*}$. For classical polymer applications, the large z regime is most interesting. $\rho(z)$ has a branch point singularity in infinity, its large z asymptotics are dominated by the factor $z^{2\nu-1}$; however, if we write $\rho(z) = z^{2\nu-1}\phi(z)$, then $\phi(z)$ is analytical in infinity and can be expanded in integer powers of $1/z$. Accordingly, the large N asymptotics of $\langle R_g^2 \rangle$ follow:

$$\langle R_g^2 \rangle_{N \gg N^*} \simeq \ell^2 N^{2\nu} A \left[1 + k_1 \left(\frac{N^*}{N} \right)^{1/2} + k_2 \left(\frac{N^*}{N} \right)^1 + \cdots \right]. \tag{2}$$

Conversely, in the region $N \ll N^*$, the approximation for $\langle R_g^2 \rangle$ is afforded by the fact that $\rho(z)$ is analytical at small z and can be expanded in integer powers of z as follows:

$$\langle R_g^2 \rangle_{1 \ll N \ll N^*} \simeq \ell^2 N \frac{A'}{12} \left[1 + k_1' \left(\frac{N}{N^*} \right)^{1/2} + k_2' \left(\frac{N}{N^*} \right)^1 + \cdots \right], \tag{3}$$

where prefactor A' should be equal to unity (which explains why we did not absorb the factor of $1/12$ into A'). Note that the latter result is an intermediate asymptotics, which means the corresponding region exists only so long as $N^* \gg 1$ is large, which means excluded volume is sufficiently small.

B. Swelling driven by topology: Crossover length

With this brief summary of results in mind we now set forward, intending to systematically compare the computational results for the behavior of trivial knots to the well-understood polymer with excluded volume.

To look at the analogy between self-avoiding polymers and trivial knots, it is useful to start [8] by identifying the crossover length for knots, an analog of N^* (1), which we call N_0. For knots, it is natural to identify the crossover value of N with the so-called characteristic length of random knotting N_0; the latter quantity is known as the characteristic length of the exponential decay of probability, $w_{triv}(N)$, of formation of a trivial knot upon random closure of the polymer ends [17]: $w_{triv} \simeq \exp(-N/N_0)$. Depending on the specifics of the model used [11,17,18], the critical length varies subtly around $N_0 \approx 300$. It is also clear qualitatively [8] and seen computationally [11] that this N_0 is about the length at which topological effect on loop swelling crosses over from marginality at $N < N_0$ to significance at $N > N_0$. In particular, it is at $N > N_0$ that the trivial knot begins to swell noticeably beyond the size of the phantom polymer [11].

C. Swelling driven by topology: Above the crossover

A number of groups reported observation of the power $\nu \approx \frac{3}{5}$ in the scaling of trivial [9–11,19] and other topologically simple [9–11] knots in the region $N > N_0$.

In Refs. [9,10,20], following the idea suggested in Ref. [22], the N dependence of $\langle R_g^2 \rangle_{triv}$ was fitted to the formula similar to Eq. (2) for self-avoiding polymers. No attempt was made at physical interpretation of the best fit values of the three coefficients (A, k_1, k_2) or the region of N where the fit was examined. In this sense, fitting with Eq. (2) was only used as an instrument to find the scaling exponent ν, which in these works was found to be strikingly consistent with the expected value of the self-avoidance exponent. A puzzling aspect of the situation is that, particularly in Ref. [10], the data was fit to Eq. (2) not only in the region $N > N_0$, but across the crossover, starting from about $N_0/3$ to about $3N_0$ (see also Ref. [20]).

At present, we are aware of no studies which provide a detailed comparison of excluded volume and trivial knotting at modest $N < N_0$. Seeking to further appraise the analogy between trivial knotting and excluded volume, in the present work, we address the two systems in the region below their respective crossovers.

D. Swelling driven by topology: Below the crossover

Formula (3) is the result of perturbation theory [15], in which conformations with overlapping segments represent a small part of conformational space and their exclusion is considered a small correction to Gaussian statistics. It is tempting to try a similar approach for knots. The idea would be to note that at small $N < N_0$, the probability of a nontrivial knot is small, which implies that restricting the loop such that it remains a trivial knot excludes only a small sector of the conformation space which, therefore, comprises a small correction to Gaussian statistics.

Let us try to imagine the realization of this idea. We want to find the swelling ratio of the trivial loop

$$\rho_{0_1} = \langle R_g^2 \rangle_{triv} / \langle R_g^2 \rangle_{phantom}. \tag{4}$$

We know that the (topology blind) ensemble average over all knots must, by definition, yield unity for the swelling ratio

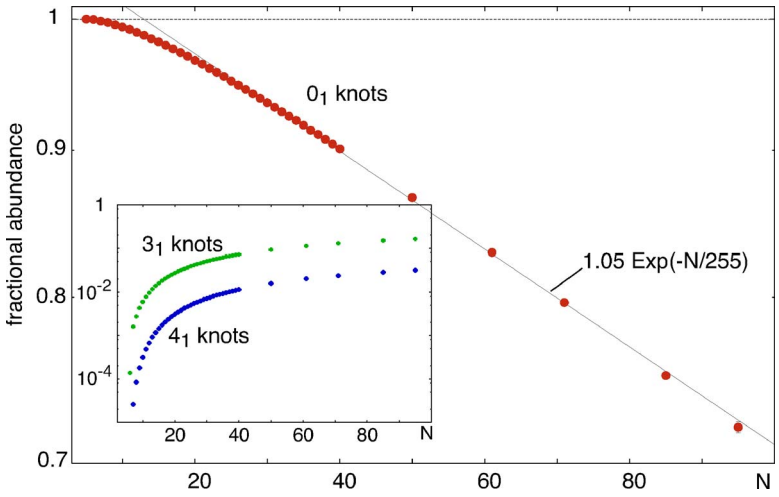

FIG. 1. (Color online) Fractional abundance of 0_1, 3_1, and 4_1 knots within the ensemble of all looped polymer chains of fixed step length. The 0_1 abundance follows the well-known [3,17,18] exponential decay, $A\exp(-N/N_0)$ with decay length $N_0=255$ and prefactor $A\approx 1.05$. Pertinent to the notion of higher-order knots acting as a perturbation is that the abundance of 3_1 and 4_1 knots, seen in the inset, is quite low in the $N \ll N_0$ region.

$$1 = P_{0_1}\rho_{0_1} + P_{3_1}\rho_{3_1} + P_{4_1}\rho_{4_1} + \cdots, \tag{5}$$

where P_i and ρ_i are, respectively, the probability and swelling ratio of the knot i. Our plan is to consider formula (5) as the equation from which we can determine the quantity of interest, ρ_{0_1},

$$\rho_{0_1} = \frac{1 - P_{3_1}\rho_{3_1} - P_{4_1}\rho_{4_1} - \cdots}{P_{0_1}}. \tag{6}$$

To this point our consideration is exact, but now we switch to hand waving arguments and guesses, mostly intuitive, but also partly justified by the simulation data. In the range of small N, the ensemble of loops consists mostly of 0_1 knots, perturbed slightly by the presence of 3_1 and an even smaller fraction of higher-order or more complex knots. We consider then N/N_0 as a small parameter: $N/N_0 \ll 1$. Of course, in the case of excluded volume, the similar limit is better justified, because N^*, Eq. (1), can, at least in principle, be arbitrarily large, leaving room for the intermediate asymptoticss $1 \ll N \ll N^*$. In the case of knots, N_0 is as large as about 300, but so far we do not know why it is large, and it seems beyond our control to make it larger. Accordingly, we cannot speak of an intermediate asymptoticss in a mathematically rigorous way [23]. Nevertheless, we assume here that the numerically large value of N_0 allows us to hope that the asymptotic argument is possible, and so we assume that N/N_0 is a small parameter, while N is still large compared at unity. We guess then that higher-order knots provide only higher-order perturbation corrections with respect to this presumably small parameter N/N_0, and we neglect their contributions, simplifying the ensemble by accounting for only 0_1 and 3_1 knots. In this case, $P_{0_1}+P_{3_1}\simeq 1$. This is justified by the data presented in Fig. 1, which shows that higher knots are very rare indeed as long as $N/N_0 \ll 1$. Since we know that $P_{0_1}\simeq\exp(-N/N_0)$, we can also find P_{3_1}. Given that we consider the $N/N_0 \ll 1$ regime, we must also linearize the exponent, which yields

$$\begin{aligned}\rho_{0_1} &\simeq \frac{1-(1-P_{0_1})\rho_{3_1}}{P_{0_1}} \\ &\simeq \frac{1-(1-e^{-N/N_0})\rho_{3_1}}{e^{-N/N_0}} \\ &\simeq (1-(N/N_0)\rho_{3_1})(1+N/N_0).\end{aligned} \tag{7}$$

The next step requires thinking about ρ_{3_1}. In principle, we can come up with a chain of equations, not unlike the Bogoliubov-Born-Green-Kirkwood-Yvon (BBGKY) chain in the theory of fluids, expressing ρ_{3_1} in terms of higher knots, etc. A more practical course is to note that for the lowest order in perturbation, with respect to the supposedly small parameter N/N_0, since ρ_{3_1} has already the small (N/N_0) coefficient in front of it, it is enough to replace ρ_{3_1} with a constant at $N/N_0 \to 0$. Thus, to the lowest order in $N/N_0 \ll 1$, we get $(N/N_0)\rho_{3_1}\simeq(N/N_0)c$, where c is a constant. We, therefore, finally obtain

$$\rho_{0_1} \simeq 1 + (N/N_0)(1-c), \tag{8}$$

or

$$\langle R_g^2\rangle_{triv} \simeq \ell^2 N\frac{1}{12}\left[1+\left(\frac{N}{N_0}\right)(1-c)\right]. \tag{9}$$

To be careful, we should mention here that simulation data for ρ_{3_1}, shown in the inset in Fig. 3, do not look constant—and they should not. To understand why we have legitimate right to replace ρ_{3_1} with a constant at $N/N_0 \to 0$, one should correctly understand the idea of intermediate asymptoticss. The point is that ρ_{3_1} does not approach 0 as N/N_0 decreases. As long as $N>1$, ρ_{3_1} approaches some nonzero constant, which we called c above. A closer look at the simulation data might allow some fit of the form $\rho_{3_1}|_{N\ll N_0} \simeq c + c'(N/N_0)^\beta$, with parameters $c'>0$ and $0<\beta<1$. A better understanding of this fit is an exciting task for the future, but for our purposes here it is enough to realize that higher resolution in the fit will only contribute to the subleading correction term in the perturbative expression for $\langle R_g^2\rangle_{triv}$, the

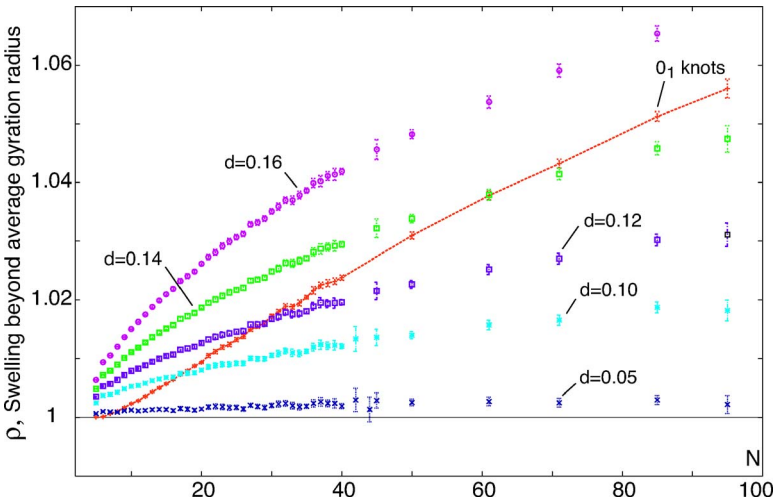

FIG. 2. (Color online) Direct comparison of excluded volume and trivial knot swelling ρ_{0_1} beyond the phantom average size for loops of fixed step length. Excluded volume is formulated in terms of N beads of diameter d, each centered at an universal joint between loop segments. Exclusion is maintained by prohibiting bead overlap, $|\vec{x}_i-\vec{x}_j|\geq d$ for all $i\neq j$. As discussed in Sec. II D, and in contrast to the region above their respective crossovers, in the small $N<N_0$ regime, trivial knots follow a functional form different from that of excluded volume loops.

leading term will remain linear in N/N_0, as stated in our formula (9). The same is true in regard to the corrections caused by the higher order knots beyond 3_1.

The difference between Eqs. (3) and (9) is immediately obvious: the former is an expansion in powers of $\sqrt{N}$, the latter starts from the first power of N. The $\sqrt{N}$ term does not occur in our expansion for knots. Note that the values of the k_i' coefficients in Eq. (3) are known [15], and this prevents the easy (and incorrect) explanation that $k_1'=0$. As regards the value of coefficient c, we do not have at present an analytical means to calculate it, we will later estimate it based on the simulation data. Thus, despite identical scaling index at large N, trivially knotted and excluded volume polymers exhibit a very different mathematical structure of N dependence in their respective gyration radii in the region of small N.

It is possible that another manifestation of the same difference is the fact that data in Ref. [10] were successfully fitted to the Eq. (2) across the crossover region, where this formula for the self-avoiding polymers is not supposed to work.

Thus, our considerations suggest that there is some fundamental difference between topology and self-avoidance in terms of their respective effects on the swelling at moderate N. In what follows, we present computational tests supporting and further developing this conclusion.

III. MODEL AND SIMULATION METHODS

We model polymer loops as a set of $N+1$ vertices, $\vec{x}_i$, embedded in $3D$, where $\vec{x}_0=\vec{x}_N$ implies loop closure. The step between successive vertices, $\vec{y}_i=\vec{x}_{i+1}-\vec{x}_i$ is constructed either from steps of fixed length, with probability density

$$P(\vec{y}_i)=\frac{1}{4\pi\ell^2}\delta(|\vec{y}_i|-\ell), \tag{10}$$

or Gaussian distributed, with probability density

$$P(\vec{y}_i)=\left(\frac{3}{2\pi\ell^2}\right)^{3/2}\exp\left(-\frac{3|\vec{y}_i|^2}{2\ell^2}\right). \tag{11}$$

Note that ℓ, the "average" step length, is defined, $\ell^2=\int P(y)y^2d^3y$. Many methods have been used to generate loops in computer simulation over the past decade. A brief review of the methods is available in Appendix A, the details of the method implemented in this work are presented in Appendix B.

Once generated, we assess the loop's size by calculating its radius of gyration

$$R_g^2=\frac{1}{2N^2}\sum_{i\neq j}|\vec{x}_i-\vec{x}_j|^2. \tag{12}$$

The mean square average radius of gyration seen over all loops is $\langle R_g^2\rangle=\frac{1}{12}(N+\beta)l^2$, where $\beta=1$ for fixed step length loops and $\beta=-1/N$ for loops of Gaussian distributed step length. Noting that the excluded volume constraint is maintained by the condition that pair distances be larger than excluded volume bead diameter, $r_{ij}=|\vec{x}_i-\vec{x}_j|, r_{ij}\geq d$, we record the minimum r_{ij} for each coil, which enables us to ascertain what maximum diameter of excluded volume d the loop corresponds to [24]. Finally, we calculate the topological state of the loop by computing the Alexander determinant, $\Delta(-1)$, and Vassiliev knot invariants of degrees 2 and 3, $v2$ and $v3$, the implementation of which is described in Ref. [25]. As the simulation progresses, averages are accumulated in a matrix, indexed over different knot types and minimum pair distances. In the end, we can collect the data to find the gyration radius for either a particular knot type irrespective of pair distances (i.e., without volume exclusion), or for a particular excluded volume value irrespective of topology.

IV. RESULTS

A. On the functional form of N dependence of the gyration radius in the moderate N regime

Figure 2 provides direct comparison of the computationally determined mean square gyration radius for trivial knots and phantom loops with excluded volume (averaged over all

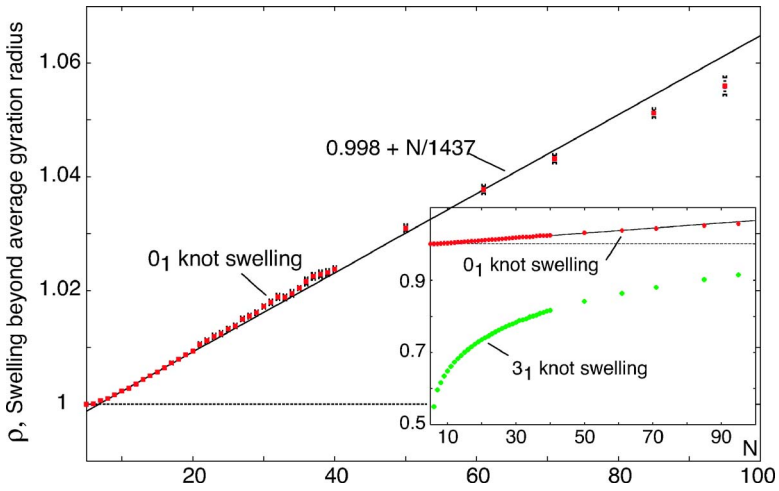

FIG. 3. (Color online) Average gyration radius data for trivially knotted loops of fixed step length. Loops were generated with the conditional probability method described in Appendix B. Swelling of the gyration radius is seen to be linear in the small N regime and can be understood initially as the result of a perturbation.

topologies), in the latter case—for various values of the bead diameter.

Note that in the figure, the gyration radius is expressed with the swelling ratio ρ, as defined in Eq. (6). The most striking feature of this figure is the differently shaped curves of swelling. The region of intermediate N visible in the figure, $1<N<N_0$, shows the plot of trivial knot swelling passing through all excluded volume curves. As seen, the very shape of the ρ_{0_1} curve is different. Specifically, all curves for the excluded volume loops are bent downward, consistent with the presence of the $\sqrt{N}$ terms in Eq. (3). In contrast, the curve for the topologically restricted trivial loop is very nicely linear. A fit of the form

$$\rho_{0_1} = 0.998 + N/1437 \approx 1 + 0.18N/N_0, \tag{13}$$

consistent our estimate, Eq. (9), where $N_0=255$, is shown in Fig. 3. Note that deviation from the linear form occurs as N increases. This is entirely expected as the crossover to asymptotic swelling of the gyration radius, $N^{2\nu}/N \sim N^{0.19}$, must occur as N grows beyond N_0.

B. Which excluded volume diameter matches most closely the topological swelling of trivial knots?

The crossover points between curves of trivially knotted loops and loops with excluded volume in Fig. 2 inspired the idea of plotting the excluded volume diameter at each N whose swelling matches the swelling of a trivial knot at the same N. As seen in Fig. 4, this mapping parameter seems to approach an asymptote at the specific diameter of $d=0.1625$. While at present it is not computationally feasible to extend the scale of N to significantly larger values, this asymptotic approach of trivial knot swelling to loops with excluded volume is consistent with the similar asymptotic swelling of $N^{2\nu}$ seen in other work [9–11].

At the same time, it is interesting to note that although the swelling parameter due to the excluded volume at $d \approx 0.16$ seems to fit the topologically driven swelling, the corresponding characteristic length N^* [see Eq. (1)] is significantly larger than N_0. To see this, we note that the excluded volume data in Fig. 2 fit reasonably well to the expression $\rho \approx 1+1.71\sqrt{N}(d/\ell)^3 = 1+\sqrt{N/N^*}$, where, therefore, $N^* = 0.34(d/\ell)^6$. Here, we determined, based on the fit, the numerical coefficient intentionally left undetermined in formula (1). At $d=0.16\ell$, we get, therefore, $N^* \approx 20\,000$, which is almost 2 orders of magnitude greater than $N_0 \approx 255$. Alternatively, this situation can be seen by finding the excluded volume diameter for which crossover length N^* matches N_0: $N^*=N_0$; the corresponding d equals $d \approx 0.33\ell$. It is fairly obvious that this value of excluded volume does not agree well with the data presented in Fig. 4. This discrepancy pos-

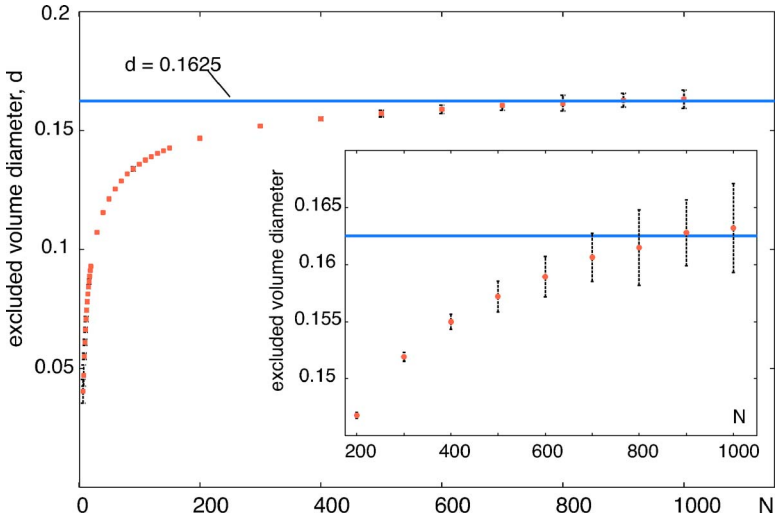

FIG. 4. (Color online) The excluded volume bead diameter which gives the same $\langle R_g^2 \rangle$ swelling as the group of trivially knotted loops. Unlike other figures in the publication, loops here are generated conditionally with Gaussian-distributed step length. This is done for feasibility reasons, as computationally, Gaussian-distributed steps are easier to generate than loops of fixed step length. As seen in the image, the excluded volume diameter seems to saturate at about $d=0.1625$. This saturation is consistent with the notion of the trivial knot gyration radius average approaching the $N^{2\nu}$ asymptotic when $N \gg N_0$. Although not tested, we expect that fixed step length loops would exhibit similar saturation at a specific excluded volume diameter.

sibly points at yet another difference between swelling driven by topology and excluded volume.

V. CONCLUSIONS

It seems quite clear from our simulation data that the analogy between excluded volume and trivial knotting does not hold at loop sizes smaller than the crossover for knots N_0. The nature of the swelling function $\rho(N)$ in this region is yet unknown. Although our cursory explanation accounts for the trivial knot data's linear trend in this regime, the similar parameter for the size of more complex knots behaves nonlinearly, and we currently have no explanation for this. A more systematic treatment of the problem is badly needed to understand the size behavior of knots.

That said, our data showing the mapping of excluded volume diameter to trivial knot size seems to reinforce the notion that asymptotically, the two classes of objects scale with the same power.

We express thanks to R. Lua of the University of Minnesota for the use of his knot analysis routines. We also wish to thank the Minnesota Supercomputing Institute for the use of their facilities. This work was supported in part by the MRSEC Program of the National Science Foundation under Grant No. DMR-0212302.

APPENDIX A: A BRIEF REVIEW OF LOOP GENERATION METHODS

A number of methods exist and have been used in the literature for the computational generation of looped polymers. The goal of generation methods is to produce statistically representative and unbiased sets of mutually uncorrelated loops. The generation of a random walk is a simple matter. Steps are chosen with isotropic probability until the desired length is reached. Creating random walks with biased probability, specifically, walks which return to the origin after a specified number of steps, is a more difficult task. As many studies of the topological properties of polymer chains have been completed, we do not intend to make an exhaustive summary of all work, but rather in broad strokes summarize the generation methods used in the field.

All methods used to generate loops can be grouped into two large categories. Methods of one group start from some loop configuration which does not pretend to be random, and then transform it in some way to randomize the set of steps making the loop. Methods of the other group build more or less random loops from the very beginning.

One of the initial techniques used for the generation of loops is the dimerization method of Chen [26,27], in which smaller sets of walks are joined end to end to form larger walks or loops. This "ring dimerization" accepts the joining of smaller walks with some probability, as self-intersections between the chains are prohibited. In addition, if the generated walk is closed to form a loop, a statistical weight is calculated to account for loop closure. Several groups have used this method [17,21], usually in the context of including excluded volume in the topological study.

Other workers [9,19] start with an initial loop conformation and then modify it by applying a number of "elbow" pivot moves on randomly selected sections of the loop. Specifically, if the loop is defined by N vertices $\{\vec{x}_i\}$, a pivot move is performed by selecting two vertices, $\vec{x}_j$ and $\vec{x}_k$, and then rotating by a random angle the intermediate vertices $\vec{x}_{j+1}$ through $\vec{x}_{k-1}$ about the axis made by $\vec{x}_k - \vec{x}_j$.

A third method in common use, the so-called "hedgehog" method [10,28], starts by generating $N/2$ pairs of mutually opposite bond vectors. The resulting set of N vectors has zero sum, and it is tempting to reshuffle them and then use them as bond vectors, thus surely obtaining a closed loop. Unfortunately, such a loop has obviously correlated segments, the most striking manifestation of which is that the loop has self-intersections with a large probability of order unity (in fact, $1/e \approx 0.37$ [29], see also a related scaling argument in Ref. [11]). To overcome this, it was suggusted by Dykhne, and later described in Ref. [28], that one imagine all N vectors plotted from the origin and thus forming something like a hedgehog, then randomly choosing pairs of vectors (hedgehog needles) and rotating the pair by a random angle about their vector sum. This operation does not change the sum of all N vectors, which remains 0, and therefore, upon sufficiently many such operations and upon reshuffling all vectors, one can hope to obtain a well-randomized loop.

The hedgehog method and elbow moves method are in fact quite similar. Indeed, in both cases, the idea is to rotate some bond vectors around their vector sum; in the hedgehog method it is done with pairs of vectors before reshuffling, in the elbow moves method it is done after reshuffling with a set of subsequent bonds, but the idea is the same. In both cases, the evolution of loop shape can be described by Rouse dynamics, known in polymer physics (see, e.g., Ref. [14]). This allows us to make a simple estimate as to how many moves are necessary in order to wash away correlations imposed by the initial loop configuration. Rouse dynamics can be understood as diffusive motion of Fourier modes. Since the longest wave Fourier mode has a wavelength which scales as N, the longest relaxation time in Rouse dynamics scales as N^2. This estimate is valid for physical dynamics in which all segments move at the same time. Translated into computational language, this implies that every monomer has to make about N^2 moves, which means that we have to make about N^3 random moves for proper removal of correlations. Unfortunately, this point is rarely mentioned in the use of these algorithms (see, however, Ref. [19]) and the number of moves between sampling is generally quite small, which puts into question the ergodicity of implementations of this algorithm.

To overcome this problem, we proposed, in Ref. [11], another method which we call the method of triangles, which does not involve any relaxation. In this method, we generate $N/3$ randomly oriented triplets of vectors with zero sum, reshuffle them, and connect them head-to-tail, thus obtaining a loop. As we shall explain in another publication, this method produces loops with insignificant correlations when N is larger than 100 or so.

Since our major attention in this paper is the range of relatively small N, we have to resort to a computationally more intensive, but reliably unbiased method based on con-

ditional probabilities. The idea is to generate step number i in the loop of N steps using the conditional probability that the given step arrives to a certain point provided that after $N-i$ more steps, the walk will arrive at the origin. This method was suggested and implemented for Gaussian chains in Ref. [30]. Here, we apply it for the loops with fixed step length.

APPENDIX B: GENERATION OF LOOPS WITH FIXED STEP LENGTH USING THE CONDITIONAL PROBABILITY METHOD

1. Derivation of the conditional probability method

A walk is composed of N steps between $N+1$ nodes, a step from nodes $\vec{x}_i$ to $\vec{x}_{i+1}$ having normalized probability, $g(\vec{x}_i,\vec{x}_{i+1},1)$. The probability for a random walk composed of N such steps is described by the Green's function which ties the steps together,

$$G(\vec{x}_0,\vec{x}_N,N) = \int g(\vec{x}_1-\vec{x}_0)g(\vec{x}_2-\vec{x}_1)\cdots g(\vec{x}_N-\vec{x}_{N-1})d\vec{x}_1 d\vec{x}_2\cdots d\vec{x}_{N-1}. \tag{B1}$$

Note that in this notation, the walk stretches from $\vec{x}_0$ to $\vec{x}_N$. The specifics of integration depend on the sort of steps which are being taken. At times, these integrations can be difficult to evaluate. In such cases, the convolution theorem can be of some utility. Suppose that the Fourier transform and inverse is defined in the usual way,

$$g_{\vec{k}} = \beta \int g(\vec{x})\exp[i\vec{k}\cdot\vec{x}]d\vec{x},$$

$$g(\vec{x}) = \beta \int g_{\vec{k}} \exp[-i\vec{k}\cdot\vec{x}]d\vec{k}. \tag{B2}$$

Note that in this formulation, $\beta=(2\pi)^{-3/2}$. The convolution theorem allows for the following expression for $N\geqslant 2$,

$$G(\vec{x}_0,\vec{x}_N,N) = (1/\beta)^{N-2} \int (g_{\vec{k}})^N \exp[-i\vec{k}\cdot(\vec{x}_N-\vec{x}_0)]d\vec{k}. \tag{B3}$$

If the step length is fixed to a certain distance ℓ, the probability distribution and its Fourier transform are expressed,

$$g(\vec{x}_0,\vec{x}_1,1)_{fixed} = \frac{\delta(|\vec{x}_1-\vec{x}_0|-\ell)}{4\pi l^2},$$

$$g_{\vec{k}} = \beta\frac{\sin(k\ell)}{k\ell}. \tag{B4}$$

Using Eqs. (B3) and (B4), along with differential volume $d\vec{k}=2\pi k^2 dk d(\cos\theta)$, the probability distribution for a walk of N fixed-length steps spanning the displacement $\vec{x}_N-\vec{x}_0$ is

$$G(\vec{x}_0,\vec{x}_N,N)_{fixed} = \beta^2 4\pi \int_0^\infty \left[\frac{\sin(k\ell)}{k\ell}\right]^N \frac{\sin(k|\vec{x}_N-\vec{x}_0|)}{k|\vec{x}_N-\vec{x}_0|}k^2 dk. \tag{B5}$$

If we use the definition of β and express sine terms as exponentials, also using $d=|\vec{x}_N-\vec{x}_0|/\ell$, then

$$G(\vec{x}_0,\vec{x}_N,N)_{fixed} = \frac{1}{2\pi^2}\int_0^\infty \frac{[\exp(ik\ell)-\exp(-ik\ell)]^N[\exp(ik\ell d)-\exp(-ik\ell d)]}{(2ik\ell)^{N+1}d}k^2 dk. \tag{B6}$$

Then using the Newton binomial $(x+y)^N=\Sigma_{m=0}^N \binom{N}{m} x^{N-m}y^m$, where $\binom{N}{m}=n!/[(n-m)!\,m!]$ yields a shiny prize, an analytically tractable expression,

$$G(\vec{x}_0,\vec{x}_N,N)_{fixed} = \frac{1}{\pi^2}\frac{1}{2^{N+2}i^{N+1}\ell^{N+1}d}\int_0^\infty \sum_{m=0}^N \binom{N}{m} \frac{[\exp(ik\ell)]^{N-m}[-\exp(-ik\ell)]^m[\exp(ik\ell d)-\exp(-ik\ell d)]}{k^{N-1}}dk. \tag{B7}$$

At this point, two further simplifications are made. The first is to extend the integration from $-\infty$ to ∞, as the integrand is even on the real axis (with proper incorporation of the factor of $\frac{1}{2}$). The second simplification is to integrate over the dimensionless number $\kappa=k\ell$. Note that the dimension of the integral remains $1/volume$.

$$G(\vec{x}_0,\vec{x}_N,N)_{fixed} = \frac{1}{\pi^2}\frac{N!}{2^{N+3}i^{N+1}\ell^3 d}\int_{-\infty}^\infty \sum_{m=0}^N \frac{(-1)^m}{(N-m)!\,m!}\frac{\exp[i\kappa(N-2m+d)]-\exp[i\kappa(N-2m-d)]}{\kappa^{N-1}}d\kappa. \tag{B8}$$

The integral which remains can be evaluated as a contour integral in the complex plane. The contour along the real axis is chosen with a small bump in the $+i$ direction at $\kappa=0$. The upper or lower arch is chosen according to Jordan's lemma. The residue at $\kappa=0$ is obtained by Taylor expanding the exponent to resolve the coefficient corresponding to the κ^{-1} term, which is the definition of a residue. The result follows:

$$\int_{-\infty}^{\infty} \frac{\exp(i\alpha\kappa)}{\kappa^{N-1}} d\kappa = \begin{cases} 0, & \text{if } \alpha \geqslant 0 \\ -2\pi i \left[\frac{1}{(N-2)!} (i\alpha)^{N-2} \right], & \text{if } \alpha < 0 \end{cases} . \tag{B9}$$

Integration winnows the sum considerably, the final result is,

$$G(\vec{x}_0, \vec{x}_N, N)_{fixed} = \frac{N(N-1)}{2^{N+2}\pi l^3 d} [J_1(N,d) - J_2(N,d)], \tag{B10}$$

where

$$J_1(N,d) = \sum_{m>(N+d)/2}^{N} \frac{(-1)^m}{(N-m)!\, m!} (N-2m+d)^{N-2}, \tag{B11}$$

and

$$J_2(N,d) = \sum_{m>(N-d)/2}^{N} \frac{(-1)^m}{(N-m)!\, m!} (N-2m-d)^{N-2}. \tag{B12}$$

A table of probabilities can then be composed. Note, however, that the probability is defined on intervals over d, listed in the right column below.

$$G(\vec{x}_0, 0, 3)_{fixed} = \frac{1}{8\pi\ell^3 d}, \quad \text{if } d \in [0,2],$$

$$G(\vec{x}_0, 0, 3)_{fixed} = \begin{cases} (1)/(8\pi\ell^3), & \text{if } d \in [0,1] \\ (3-d)/(16\pi d\ell^3), & \text{if } d \in [1,3] \end{cases},$$

$$G(\vec{x}_0, 0, 4)_{fixed} = \begin{cases} (8-3d)/(64\pi\ell^3), & \text{if } d \in [0,2] \\ (d-4)^2/(64\pi\ell^3 d), & \text{if } d \in [2,4] \end{cases},$$

$$G(\vec{x}_0, 0, 5)_{fixed} = \begin{cases} (5-d^2)/(64\pi\ell^3), & \text{if } d \in [0,1] \\ (2d^3 - 15d^2 + 30d - 5)/(192\pi\ell^3 d), & \text{if } d \in [1,3] \\ -(d-5)^3/(384\pi\ell^3 d), & \text{if } d \in [3,5] \end{cases},$$

$$G(\vec{x}_0, 0, 6)_{fixed} = \begin{cases} (5d^3 - 24d^2 + 96)/(1536\pi\ell^3), & \text{if } d \in [0,2] \\ (-5d^4 + 72d^3 - 360d^2 + 672d - 240)/(3072\pi\ell^3 d), & \text{if } d \in [2,4] \\ (d-6)^4/(3072\pi\ell^3 d), & \text{if } d \in [4,6] \end{cases} . \tag{B13}$$

These piecewise-defined probability distributions approach the shape of the corresponding quantity for Gaussian-distributed step length,

$$G(\vec{x}_0, \vec{x}_N, N)_{gaussian} = \left(\frac{3}{2\pi N\ell^2} \right)^{3/2} \exp\left[-\frac{3}{2N\ell^2} (\vec{x}_N - \vec{x}_0)^2 \right]. \tag{B14}$$

Due to the complexity and computational expense of the conditional method, and noting the apparent similarity of the two curves, one might be tempted to substitute the Gaussian formulation, Eq. (B14), when N is above some threshold $N > N_c$. Our own experience with this approximation leads us to discourage the intermingling of the two distributions. When included, at even the large $N_c = 30$, a sharp discontinuity in the curve of curve for ρ_{0_1} vs N (Fig. 3) was visible at N_c. We hypothesize that substitution of the Gaussian formulation, Eq. (B14), for the fixed-step formulation, Eq. (B10), allows for slightly more inflated loop conformations and thus leads to a discontinuity when the approximation is used in the simulation code at $N > N_c$.

2. Implementation of conditional probability method

Generation of a random walk which is looped, i.e., $\vec{x}_N - \vec{x}_0 = 0$, can be achieved with the use of the already derived equations. Imagine that a walk of $N+M$ steps is underway and M steps have already been taken. This means that a walk of N steps remains, which starts at the present location $\vec{x}_0$ and finishes at the starting point $\vec{x}_N$. The probability distribution for the next step, from $\vec{x}_0$ to $\vec{x}_1$, can then be written,

$$P(\vec{x}_0 | \vec{x}_1) = \frac{G(\vec{x}_0, \vec{x}_1, 1) G(\vec{x}_1, \vec{x}_N, N-1)}{G(\vec{x}_0, \vec{x}_N, N)}. \tag{B15}$$

In principle, one could generate new steps with probability isotropic in direction, accepting them with conditional probability defined by Eqs. (B10) and (B15) or (B14). In the interest of efficiency, a better method is to generate random steps within these probability distributions. Now discussed is the way to transform a flat random distribution [that produced by the UNIX math function DRAND48(), for example] into the distribution above. If the flatly distributed variable is q, i.e., $P(q) = 1$ on $[0,1)$, 0 elsewhere, the following equation,

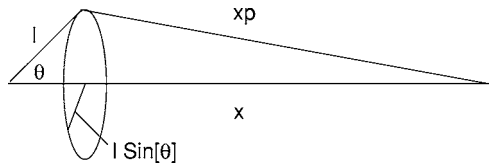

FIG. 5. This geometry is used in the implementation of the conditional probability loop generation method.

with $\vec{d}=\vec{x}_N-\vec{x}_1$, defines the transform to the conditional distribution above, $G(\vec{x})$,

$$\int_0^q P(q')d(q') = \int_0^{f(q)} \frac{G(\vec{x}_0,\vec{x}_1,1)G(\vec{d},0,N-1)}{G(\vec{x}_0,\vec{x}_N,N)} d(\vec{d}). \tag{B16}$$

In this statement of normalization, the function of importance is $f(q)$, which defines the way the two probability distributions are made equal.

In principle, the problem is now solved. A complete set of probability distributions for walks of fixed or Gaussian step length has been defined, and the the formula which maps that distribution to a flat, machine-generated distribution has also been expressed. If the form of Eq. (B16) is simple enough, meaning relatively small N, the integral equation can be solved directly for $f(q)$. In practice however, $N>5$ is an interesting regime and a different technique must be used to obtain $f(q)$.

For the case of finishing a random walk of fixed length steps ℓ, which is $\vec{x}$ away from the ending point, and has N steps alloted to get to that point, we use the geometry shown in Fig. 5. In this diagram, $\vec{x}_p$ is the new distance away from the endpoint after the present step is taken. Thus, the expression above becomes

$$\int_0^q P(q')dq' = \int_0^{f(q)} \frac{G(\vec{l},1)G(\vec{x}_p,N-1)}{G(\vec{x},N)} d(\vec{x}_p), \tag{B17}$$

where, for convenience, the following syntax is used, $G(\vec{b},0,N)=G(\vec{b},N)$.

Of course, the single step $G(\vec{d},1)$ is a delta function, $\delta(|\vec{d}|-\ell)/4\pi\ell^2$, so the integration over $d(\vec{x}_p)$ occurs over most or all of the spherical shell created by the possible orientations of ℓ. Integration over the shell (about the axis made by $\vec{x}$) is performed in "rings," each ring having circumference $2\pi\ell \sin[\theta]$, and width $\ell d(\theta)$, with resulting differential area, $dA=2\pi\ell^2 \sin[\theta]d(\theta)$. θ is integrated over the range $[0,\pi]$.

It should be apparent that ${x_p}^2=x^2+\ell^2+2x\ell\cos[\theta]$. This yields the differential transform, $\sin[\theta]d(\theta)=(x_p/x\ell)d(x_p)$. This simplification allows the integration of Eq. (B17) in the following way:

$$\int_0^q P(q')dq' = \frac{1}{2\ell x G(x,N)} \int_{min}^{f(q)} G(\vec{x}_p,N-1)x_p d(x_p). \tag{B18}$$

This expression is normalized to 1 if integrated over appropriate x_p bounds. In most cases, those bounds are $[x-\ell,x+\ell]$, although the physical limit on the upper bound, $x_p \leqslant (N-1)\ell$ is necessary to keep the walk from straying too far from the origin. Additionally, if the walk is very close to the origin $x<\ell$, the integration bounds, $[\ell+x,\ell-x]$, are used.

As Eq. (B12) for fixed step length probability is defined as a polynomial, integration of that polynomial, described by Eq. (B18), can be performed exactly within simulation computer code, and the resulting equation for $f(q)$ solved numerically. In practice, we use the Gnu Multiple Precision library to represent the polynomial coefficients and values as *rational* numbers. From a computational standpoint, this is significantly more expensive than representing coefficients as double floating points, but using rationals allows us to represent all outputs of the polynomial with great accuracy, which is the goal of this simulation method. At a relatively small number of steps, the coefficients become quite small, for example, at $N=15$, in the region $x\in[13,15]$, Eq. (B12) reads $-(d-15)^{13}/[40\,809\,403\,514\,880(\ell^3 d\pi)]$. We feel the need in this routine to retain accuracy when performing operations such as $P-Q$, where $P\gg 1$ and $Q\gg 1$, but $(P-Q) \ll P,Q$. In order to retain the accuracy of the conditional formulation, it was imperative to perform this rational number algebra. For the interested reader, we provide a table of these polynomial coefficients as supplementary materials [31].

[1] Lord Kelvin, Trans. - R. Soc. Edinbrgh **25**, 217 (1868).
[2] F. Dean, A. Stasiak, T. Koller, and N. Cozzarelli, J. Biol. Chem. **260**, 4975 (1985).
[3] V. Rybenkov, N. Cozzarelli, and A. Vologodskii, Proc. Natl. Acad. Sci. U.S.A. **90**, 5307 (1993).
[4] Y. Arai, R. Yasuda, K. Akashi, Y. Harada, H. Miyata, K. Kinosita, and H. Itoh, Nature (London) **399**, 446 (1999).
[5] X. R. Bao, H. J. Lee, and S. R. Quake, Phys. Rev. Lett. **91**, 265506 (2003).
[6] J. des Cloizeaux, J. Phys. (Paris), Lett. **42**, L433, (1981).
[7] S. R. Quake, Phys. Rev. Lett. **73**, 3317 (1994).
[8] A. Y. Grosberg, Phys. Rev. Lett. **85**, 3858 (2000).
[9] H. Matsuda, A. Yao, H. Tsukahara, T. Deguchi, K. Furuta, and T. Inami, Phys. Rev. E **68**, 011102 (2003).
[10] A. Dobay, J. Dubochet, K. Millett, P. Sottas, and A. Stasiak, Proc. Natl. Acad. Sci. U.S.A. **100**, 5611 (2003).
[11] N. T. Moore, R. Lua, and A. Grosberg, Proc. Natl. Acad. Sci. U.S.A. **101**, 13431 (2004).
[12] Paul J. Flory, J. Chem. Phys. **17**, 303 (1949).
[13] N. Madras and G. Slade *Self-Avoiding Walk* (Birkhuser, Boston, 1999).
[14] A. Y. Grosberg and A. R. Khokhlov, *Statistical Physics of*

Macromolecules (AIP, Melville, NY, 1994) p. 91.
[15] Hiromi Yamakawa, *Modern Theory of Polymer Solutions* (Harper and Row, New York, 1971).
[16] E. F. Casassa, J. Polym. Sci. A **3**, 605 (1965).
[17] K. Koniaris and M. Muthukumar, Phys. Rev. Lett. **66**, 2211 (1991).
[18] T. Deguchi and K. Tsurusaki, Phys. Rev. E **55**, 6245 (1997).
[19] J. M. Deutsch, Phys. Rev. E **59**, R2539 (1999).
[20] K. Millett, A. Dobay, and A. Stasiak, Macromolecules **38**, 601 (2005).
[21] M. K. Shimamura and T. Deguchi, Phys. Rev. E **65**, 051802 (2002).
[22] E. Orlandini, M. Tesi, E. J. van Rensburg, and S. Whittington, J. Phys. A **31**, 5953 (1998).
[23] G. I. Barenblat, *Scaling, Self-Similarity and Intermediate Asymptotics* (Cambridge University Press, Cambridge, U.K., 1996).
[24] Use of this computational trick means that ensembles for similarly excluded volume diameters suffer from significant correlation. Nevertheless, given the qualitative nature of our investigation of excluded volume, we feel this is not a dangerous correlation to include.
[25] R. Lua, A. Borovinskiy, and A. Grosberg, Polymer **45**, 717 (2004).
[26] Y. Chen, J. Chem. Phys. **74**, 2034 (1981).
[27] Y. Chen, J. Chem. Phys. **75**, 5160 (1981).
[28] K. Klenin, A. Vologodskii, V. Anshelevich, A. Dykhne, and M. Frank-Kamenetskii, J. Biomol. Struct. Dyn. **5**, 1173 (1988).
[29] P. Flajolet and M. Noy, in *Formal Power Series and Algebraic Combinatorics*, edited by A. V. Mikhalev, D. Krob, and A. A. Mikhalev (Springer, New York, 2000), pp. 191–201.
[30] A. V. Vologodskii and M. D. Frank-Kamenetskii, Sov. Phys. Usp. **134**, 641 (1981).
[31] See EPAPS Document No. E-PLEEE8-72-106512 for (brief description). This document can be reached via a direct link in the online article's HTML reference section or via the EPAPS homepage (http://www.aip.org/pubservs/epaps.html).