# Ranking Knots of Random, Globular Polymer Rings

M. Baiesi,[1] E. Orlandini,[2,3] and A. L. Stella[2,3]

[1]*Dipartimento di Fisica, Università di Firenze, and Sezione INFN, Firenze, I-50019 Sesto Fiorentino, Italy*
[2]*Dipartimento di Fisica and Sezione CNR-INFM, Università di Padova, I-35131 Padova, Italy*
[3]*Sezione INFN, Università di Padova, I-35131 Padova, Italy*

(Received 29 March 2007; published 1 August 2007)

An analysis of extensive simulations of interacting self-avoiding polygons on cubic lattice shows that the frequencies of different knots realized in a random, collapsed polymer ring decrease as a negative power of the ranking order, and suggests that the total number of different knots realized grows exponentially with the chain length. Relative frequencies of specific knots converge to definite values because the free energy per monomer, and its leading finite size corrections, do not depend on the ring topology, while a subleading correction only depends on the crossing number of the knots.

DOI: 10.1103/PhysRevLett.99.058301 PACS numbers: 36.20.Ey, 02.10.Kn, 87.15.Aa, 64.60.Ak

Issues related to the probability of realization of configurations with specific knots in closed random chains play a major role in topological polymer statistics [1,2] and in its applications to macromolecular and biological physics [3]. Interest in the spectrum of different knots realized in random polymer models is stimulated, e.g., by the need for comparison with the circular DNA extracted from some viral capsids [4,5], in the hope of identifying specific biological mechanisms of knot formation. The low number of globular proteins for which a knot has been detected in the native state [6,7] marks a striking difference with respect to the general collapsed phase of homopolymers, in which a definitely higher knotting frequency is expected. Understanding the reasons for such a difference is certainly a key issue for the formulation of adequate statistical models of proteins [8]. The interest in collapsed polymers is also stimulated by the recent realization that their knots, unlike prime knots in the good solvent case, are on average completely delocalized along the backbone [9]. Understanding if and up to what extent topological invariants can affect the globular state in such conditions is an intriguing fundamental issue. Already in the swollen regime, the problem of precisely determining the possible dependence on topology of the free energy per monomer and of the exponent specifying its correction $\propto \ln N/N$ in the limit where the number $N$ of monomers approaches infinity, remains open [10,11]. This in spite of the fact that the localized character of prime knots should represent a simplifying feature. In the globular state, similar issues have never been addressed, and their discussion should include the free energy correction associated with the existence of the globule-solvent interface [12,13].

In the present Letter, we investigate the different topologies realized by the equilibrium configurations of a collapsed ring polymer. In spite of the considerable complexity of these configurations, we show that a relatively simple statistical law governs the frequencies of the various knots, with far reaching consequences. On one hand, it allows us to argue the rate at which the amplitude of the spectrum of different knot topologies grows with increasing ring length. At the same time, the emerging scenario confirms that knots are delocalized and clarifies how topology controls the statistics of the globular state.

A cornerstone in topological polymer statistics has been the realization that, for self-avoiding polygons (SAPs), unknotted configurations are entropically disfavored, so that their probability approaches zero exponentially with growing chain length [11,14,15]. However, after this important step, progress in the statistical analysis of knot complexity was hindered by the circumstance that unknotted configurations in swollen, good solvent regimes remain overwhelmingly dominant even for relatively very long chains. On the other hand, for polymer rings in bad solvent, the collapsed state leads us to expect a much higher probability of knotting compared to the swollen case with the same chain length. This circumstance suggests collapsed polymers as ideal systems for the study of how topological complexity develops and statistically distributes itself with growing chain length.

As a model of the large scale behavior of a long flexible polymer in a solvent, we adopt the self-avoiding walk on cubic lattice. Attractive energies ($\varepsilon = -1$) between nonconsecutive nearest neighbor visited sites allow us to work in the collapsed state at $T < T_\Theta$, where $T_\Theta$ is the theta-collapse temperature [16]. In such a regime, it is rather difficult to sample a sufficient number of uncorrelated polymer configurations. We use the pruned enriched Rosenbluth method [17] (PERM), which was successfully used also for the search of native states of protein models [18]. Our computational problem is made heavier by the fact that the closed chain configurations, i.e., the SAPs generated by PERM, are only a small fraction of the total. A more serious difficulty is the topological analysis of the knot type present in a given closed chain configuration. The configurations of long, collapsed SAPs are very intricate geometrically, and their projections on planes present a huge number of crossings. This makes impossible the calculation of the topological invariants necessary for the knot identification. In order to circumvent this difficulty, we simplify each sampled configuration before per-

 

forming the analysis of invariants. To this purpose, we apply to the configuration a smoothing algorithm, which progressively reduces the length of the chain, while keeping its topology unaltered (for a similar procedure, see [5]). This algorithm is in fact a grand-canonical simulation [19] in which the step fugacity is fixed low enough to cause a rapid reduction of step number in the SAP to be analyzed. Projections of such shrunk SAP are then analyzed by the HOMFLY polynomial in the "Knotscape" program [20]. Our code allows us to resolve prime knots up to 11 essential crossings and composite knots up to 5 components.

Along the above lines, we perform a systematic analysis of the knot spectrum for $N$-step SAPs at $T = 2.5 < T_\Theta$ ($T_\Theta = 3.595$ [21]), for chain lengths up to $N = 1800$. A first result concerns the behavior of the probability $P_\varnothing(N)$ of configurations with knot $k = \varnothing$, where $\varnothing$ indicates the unknot, as a function of $N$. Our procedure allows the classification of almost all configurations for $N$ up to 800, while from $N = 1000$ to $N = 1800$, we have a progressive degradation of performances, reaching a 59% of unresolved configuration for $N = 1800$. There is thus an uncertainty on the normalization needed to calculate probabilities like $P_\varnothing(N)$. The most plausible scenario is that unresolved configurations have complex knots, especially prime knots with more than 11 crossings or composite ones with more than 5 prime components. They enter thus in the statistics as "unresolved." We see an exponential decay $P_\varnothing \sim e^{-N/N_0}$, for increasing $N$. The decay constant is $N_0 \simeq 420$ (see Fig. 1). This value is 2 orders of magnitude smaller than that expected for $T = \infty$ ($N_0 \approx 2.4 \times 10^5$ [14,15]), when SAP statistics is controlled by excluded volume alone. This small $N_0$ indicates that configurations with knots have appreciable probability already for relatively short SAPs. In a recent work on Hamiltonian loops on the cubic lattice [22], a value of $N_0 \lesssim 196$ was estimated, about half the one we determine at $T = 2.5$. If we think of Hamiltonian loops as a ($T = 0$)-like situation, a lower value of $N_0$ should indeed be expected if the trend of $N_0$ decreasing with $T$ is general.

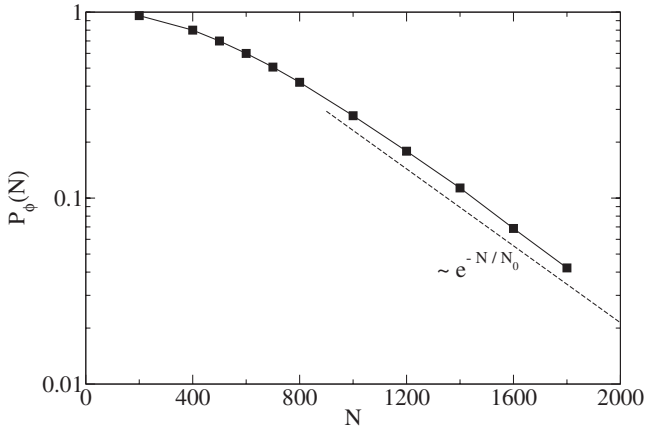


FIG. 1. Probability of finding an unknot. The straight dashed line represents the exponential fit with $N_0 = 420$.

A further step was the analysis of the knot type $k$ of each sampled configuration and of its probability $P_k(N)$ in the statistics. We rank in decreasing order these probabilities, and, by simply indicating as $P_q$ the probability of the knot $k$ with rank $q$ ($q = 1, 2, 3, \ldots$), we obtain the log-log plots reported in Fig. 2. These plots correspond to increasing numbers of sampled configurations for $N = 600$. All curves display the same slope and overlap in the first part. They only differ for the cutoff: the richer the sample, the larger the maximum rank of the realized knots. At the same time, upon varying sample size, we observe stability of the rank ordering of the knots corresponding to the initial linear parts of the plots. Thus, the estimates of $P_k(600)$ should not be affected by systematic errors. All this means that the observed cutoff in rank is only due to limited sampling and has nothing to do with the presumably much higher cutoff on the spectrum of different realizable knots due to the fact that the SAP length is finite. The power-law behavior shown by all the plots is therefore a robust feature of the data in this regime.

Figure 3 shows similar plots, this time for varying $N$ and with quite rich samples. For clarity, the effects of the statistical cutoffs are not shown, and probabilities are divided by the unknot probability $P_\varnothing(N)$. The slope of the plots is an increasing function of $N$. The estimates of the slope extrapolate to a value $-0.61(3)$ (see Fig. 3) for $N \to \infty$. Thus, for large $N$, the rank ordering statistics of the knots obey a law of the Zipf type:

$$P_q(\infty)/P_\varnothing(\infty) \sim q^{-r} \tag{1}$$

with $r \approx 0.6$. The Zipf law was first observed in the context of linguistics [23], where it rules the rank ordering in frequency of different words in a text. Since then, it has emerged in different fields, ranging from economics to disordered systems [24]. For sure, the validity of such simple law for knots in random polymers reveals a remark-

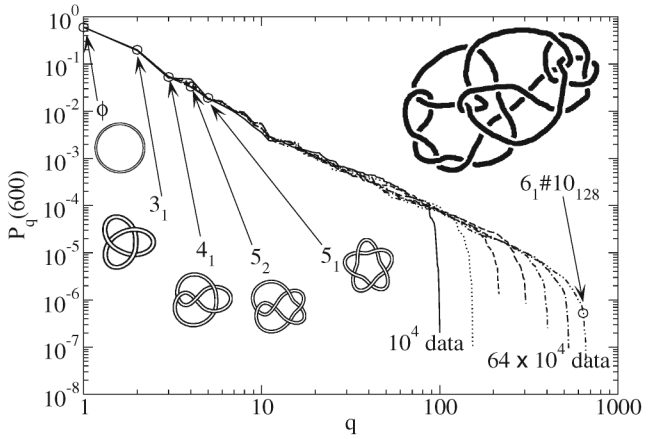


FIG. 2. Probability of rank-ordered knots for $N = 600$. Curves are for sample sizes increasing as a power of 2, from $10^4$ to $64 \times 10^4$. Some simple knot topologies (unknot, $3_1$, $4_1$, $5_2$, and $5_1$) and a more complicated one ($6_1\#10_{128}$) are also associated with their ranking.

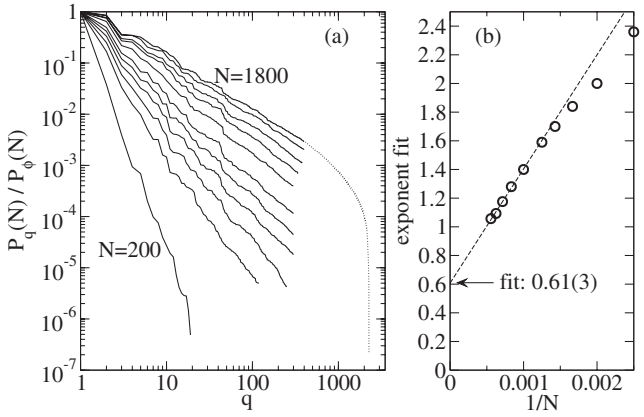


FIG. 3. (a) Power-law ranges of $P_q(N)/P_\varnothing(N)$, for $N = 200$, 400, 500, 600, 700, 800, 1000, 1200, 1400, 1600, and 1800. For $N = 1800$ also, the cutoff, due to limited sampling, is shown. (b) Extrapolation of the exponent of the Zipf law for $N \to \infty$.

able and unsuspected degree of organization of such form of topological complexity.

The asymptotic validity of Eq. (1) with exponent $r < 1$ allows us to infer a fundamental property of the knot spectrum, which at first sight would seem inaccessible to any numerical investigation. As mentioned above, there must be a cutoff $q_{\max}(N)$ for the Zipf law (1). Even if the statistical cutoff due to finite sampling occurs at much lower $q$'s, the dependence of $q_{\max}(N)$ on $N$ is a key information on the spectrum, providing a lower bound for the maximum number of different knot topologies a SAP of length $N$ can host. Assuming validity of the Zipf behavior in Eq. (1) up to $q = q_{\max}(N)$, in view of the normalization condition $\sum_q P_q(N) = 1$ valid for any $N$, one easily concludes that

$$1 + \sum_{q=2}^{q_{\max}(N)} q^{-r} \sim q_{\max}^{1-r} \sim P_\varnothing(N)^{-1}.$$

Thus, $q_{\max}(N) \sim P_\varnothing(N)^{-1/(1-r)} \sim e^{N/0.4N_0}$. This means that the number of different possible topologies a SAP can host grows (at least) exponentially with $N$.

The validity of the law in Eq. (1) also for the asymptotic ranking of knots is induced by some balance in the knot frequencies, which should not diverge relative to each other for $N \to \infty$. For a collapsed globule in equilibrium, theoretical arguments and numerical results suggest the following form for the large $N$ behavior of the canonical partition function [12]:

$$e^{-F/kT} \sim A e^{\kappa N} e^{\kappa_1 N^{2/3}} N^{\alpha-2} \qquad (2)$$

where $F$ is the free energy, $\kappa$ and $\kappa_1$ are reduced free energies per monomer and per interface monomer, respectively, and $\alpha$ is an exponent [16]. The stretched exponential factor containing $\kappa_1$ clearly implies an interfacial contribution to the free energy since the area of the globule-solvent interface is expected to grow $\sim N^{2/3}$. It is natural to expect an asymptotic behavior of the form in Eq. (2) also for ensembles with fixed knot topology like those considered here. The value of $\alpha$ could depend on the topology of the globule in such ensembles. Indeed, in the $T \to \infty$ case, there are indications that the analog of $\alpha$ for prime knots differs sensibly from that for composite knots [10] (as confirmed in [11]). Dependences on topology could not be excluded, *a priori*, also for parameters like $\kappa$ and $\kappa_1$. Indeed, the equivalent of $\kappa$ in the $T \to \infty$ limit for the unknotted ring is rigorously known to be different from the $\kappa$ in the ensemble with unrestricted topology [2]. On the other hand, the $\kappa$ of the unknotted ring has been conjectured to be the same as that of any other knotted ring for noninteracting SAPs [10].

An analysis of ratios $P_k/P_\varnothing$ should reveal the possible dependence of $\kappa$, $\kappa_1$, $\alpha$, and $A$ on the knot type. These ratios are shown in Fig. 4 as a function of $1/N$ for the simplest prime and composite knots [25]. They do not diverge. This implies that for the analyzed different knots in the collapsed polymer ring, the parameters $\kappa$, $\kappa_1$, and $\alpha$ must be the same. In particular, unlike in the $T \to \infty$ case, the entropic $\alpha$ exponent should be the same for all knots. On the other hand, Fig. 4 reveals that one should include a knot-dependent subleading factor $\exp(-\delta_k/N)$ in the form of the partition function (2). The differences $\delta_k - \delta_\varnothing$ are proportional to the slopes of data sets in log-scale in Fig. 4. These have values that essentially are determined by the crossing number $n_c$ of the knots (see Fig. 5). For example, $\delta_{3_1\#3_1} \approx \delta_{6_1}$, and so on. A simple power-law increase $\delta_k - \delta_\varnothing \sim n_c^{1.45}$ fits rather well the data in Fig. 5.

We stress that for swollen polymers, there are clear numerical evidences [10,11] of the conjecture that $\alpha_k = \alpha_\varnothing + \pi_k$, where $\pi_k$ is the number of prime components of the knot. This can be explained by taking into account that each such component is localized along the chain [9–11], bringing an entropic factor $\sim N$ to the partition function. On the other hand, the convergence of the relative frequen-

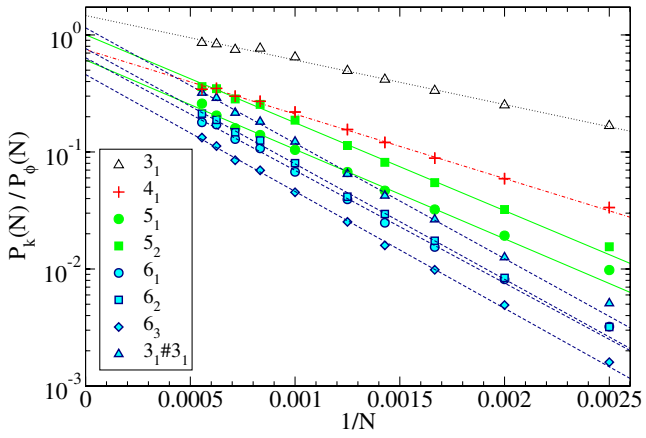


FIG. 4 (color online). Probability of a knot type $k$ over the probability of an unknot, vs $1/N$, for some knots. Each ratio converges towards an asymptotic value $A_k/A_\varnothing$. Straight lines are fits $\sim \exp[-(\delta_k - \delta_\varnothing)/N]$.

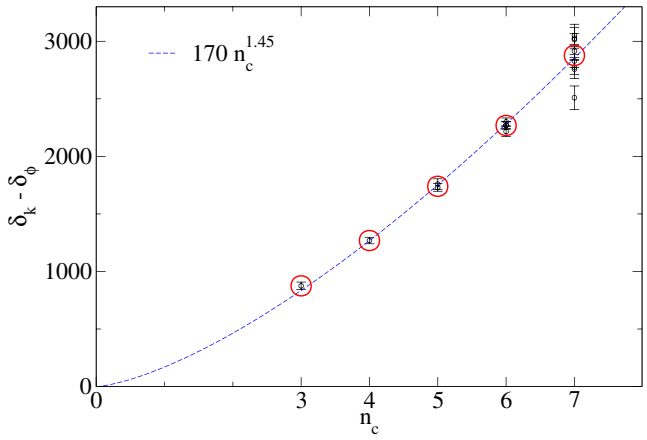


FIG. 5 (color online). Fitted $\delta_k - \delta_\varnothing$ as a function of the minimal crossing number $n_c$ for knots $3_1$ ($n_c = 3$), $4_1$ ($n_c = 4$), $5_1$ and $5_2$ ($n_c = 5$), $3_1\#3_1$, $6_1$, $6_2$, and $6_3$ ($n_c = 6$), and for all knots with $n_c = 7$. (Large circles denote averages for each $n_c$.) $\delta_k$'s are grouped in bands corresponding to $n_c$'s, i.e., to a good approximation, they depend only on the number of essential crossings of a knot type. A fit is also shown, suggesting that $\delta_k - \delta_\varnothing$ grows as a power of $n_c$.

cies of knots to definite values rules out this picture for collapsed polymers ($\alpha_k = \alpha_\varnothing$).

In summary, we have shown that a great deal of information can be gained from a numerical investigation of knots in collapsed polymers. The study of ratios of knot frequencies and the established Zipf type of law for the ranking in frequency of the knots allow us to draw very solid conclusions concerning the universality with respect to topology of several statistical parameters characterizing the globules. We find that the only nonuniversal parts of the knot frequencies relative to a reference frequency (the one of the unknotted rings) are an asymptotic ($N \to \infty$) knot-dependent amplitude ratio and the rate of convergence to this ratio, which seems to be determined only by the crossing number of the knots. These results are consistent with the expectation that knots are delocalized in the collapsed regime [9]. The Zipf law also enables us to predict how the spectrum of different topologies grows in amplitude with growing ring length.

We thank A. Stasiak for discussions. This work was supported by FIRB01 and MIUR-PRIN05. M. B. acknowledges financial support from EC FP6 project "EMBIO" (EC Contract No. 012835).

[1] H. L. Frisch and E. Wasserman, J. Am. Chem. Soc. **83**, 3789 (1961); M. Delbruck, *Mathematical Problems in the Biological Sciences*, edited by R. E. Bellman, Proc. of Symp. in Appl. Math. Vol. 14 (Am. Math. Soc., Providence, Rhode Island, 1962), p. 55; S. F. Edwards, Proc. Phys. Soc. London **91**, 513 (1967); M. D. Frank-Kamenetskii, A. V. Lukashin, and A. V. Vologodskii, Nature (London) **258**, 398 (1975).

[2] D. W. Sumners and S. G. Whittington, J. Phys. A **21**, 1689 (1988).

[3] V. V. Rybenkov, N. R. Cozzarelli, and A. V. Vologodskii, Proc. Natl. Acad. Sci. U.S.A. **90**, 5307 (1993); S. Y. Shaw and J. C. Wang, Science **260**, 533 (1993); A. Stasiak, V. Katritch, J. Bednar, D. Michoud, and J. Dubochet, Nature (London) **384**, 122 (1996).

[4] J. Arsuaga, M. Vazquez, S. Trigueros, D. W. Sumners, and J. Roca, Proc. Natl. Acad. Sci. U.S.A. **99**, 5373 (2002); J. Arsuaga *et al.*, Proc. Natl. Acad. Sci. U.S.A. **102**, 9165 (2005).

[5] C. Micheletti, D. Marenduzzo, E. Orlandini, and D. W. Sumners, J. Chem. Phys. **124**, 064903 (2006).

[6] W. R. Taylor, Nature (London) **406**, 916 (2000).

[7] P. Virnau, L. A. Mirny, and M. Kardar, PLOS Comp. Biol. **2**, e122 (2006).

[8] R. C. Lua and A. Y. Grosberg, PLOS Comp. Biol. **2**, e45 (2006).

[9] B. Marcone, E. Orlandini, A. L. Stella, and F. Zonta, J. Phys. A **38**, L15 (2005); Phys. Rev. E **75**, 041105 (2007); see also: E. Orlandini, A. L. Stella, and C. Vanderzande, Phys. Rev. E **68**, 031804 (2003); A. Hanke, R. Metzler, P. G. Dommersnes, Y. Kantor, and M. Kardar, Eur. Phys. J. E **12**, 347 (2003).

[10] E. Orlandini, M. C. Tesi, E. J. Janse van Rensburg, and S. G. Whittington, J. Phys. A **31**, 5953 (1998).

[11] M. Baiesi, E. Orlandini, and A. L. Stella (to be published).

[12] A. L. Owczarek, T. Prellberg, and R. Brak, Phys. Rev. Lett. **70**, 951 (1993).

[13] M. Baiesi, E. Orlandini, and A. L. Stella, Phys. Rev. Lett. **96**, 040602 (2006).

[14] E. J. Janse van Rensburg, *Physical Knots: Knotting, Linking, and Folding Geometric Objects in* $R^3$, edited by J. A. Calvo *et al.*, Contemp. Math. Vol. 304 (Amer. Math. Soc., Rhode Island, 2002), p. 125.

[15] A. Yao, H. Matsuda, H. Tsukahara, M. K. Shimamura, and T. Deguchi, J. Phys. A **34**, 7563 (2001).

[16] C. Vanderzande, *Lattice Models of Polymers* (Cambridge University Press, Cambridge, 1998).

[17] P. Grassberger, Phys. Rev. E **56**, 3682 (1997).

[18] H. P. Hsu, V. Mehra, W. Nadler, and P. Grassberger, J. Chem. Phys. **118**, 444 (2003).

[19] B. Berg and D. Foerster, Phys. Lett. B **106**, 323 (1981); C. Aragão de Carvalho, S. Caracciolo, and J. Fröhlich, Nucl. Phys. B **215**, 209 (1983).

[20] http://www.math.utk.edu/~morwen/knotscape.html.

[21] M. C. Tesi, E. J. Janse van Rensburg, E. Orlandini, and S. G. Whittington, J. Phys. A **29**, 2451 (1996).

[22] R. Lua, A. L. Borovinskiy, and A. Y. Grosberg, Polymer **45**, 717 (2004).

[23] G. K. Zipf, *Human Behaviour and the Principle of Least-Effort* (Addison-Wesley, Cambridge, MA, 1949).

[24] See, for example, the web p. http://linkage.rockefeller.edu/wli/zipf/index_ru.html.

[25] Note that asymptotically $3_1$ becomes slightly more frequent than $\varnothing$. The unknot however remains the most frequent knot if topologies are distinguished also according to their chiralities, such that the two mirror versions $3_1^+$ and $3_1^-$ represent two different knots.