Available online at www.sciencedirect.com
ScienceDirect
Physics Procedia 6 (2010) 117–125
Physics Procedia
www.elsevier.com/locate/procedia

Detection and visualization of physical knots in macromolecules

Peter Virnau

Johannes Gutenberg-Universität, Staudinger Weg 7, 55099 Mainz, Germany

Abstract

This manuscript provides a pedagogical introduction on how to determine and visualize simple physical knots occurring in polymers, proteins and DNA. We explain how the Alexander polynomial is computed and implemented in a simulation code, and how the structure can be simplified beforehand to save computer time. The concept of knottedness can also be extended in a statistical framework to chains which are not closed. The latter is exemplified by comparing statistics of knots in open random walks and closed random loops.
Keywords: keyword:, knots, Alexander polynomial, visualization, polymers, proteins,

1. Introduction

Apart from their apparent practical value, knots have captured the imagination of researchers from all branches of science for more than 140 years. Before knot theory became part of mathematical topology, William Thompson (who later became Lord Kelvin) suggested that "vortex knots" may provide a model for describing the diversity of atoms [1]. These studies motivated other mathematical physicists like Maxwell and Tait to work on this topic as well and eventually resulted in the first classification schemes for knots [2]. Even though Thompson's beautiful hypothesis did not come true, these efforts nevertheless laid the foundations on which modern mathematical knot theory emerged.

In the second half of the 20th century, interest in knots was revitalized in the physical sciences [3]. Delbrück, Wassermann and Frisch conjectured in the early 60s that all polymers will eventually become knotted if they are long enough [4, 5]. This statement is interesting - not only because knots are not included in the standard theories of polymers. They also provide a measure of self-entanglement and influence the dynamics and material properties of the chain. Since then, computer simulation studies [6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16] have contributed considerably to a better understanding of knottedness in real systems, and small knotted molecules have even been synthesized [17, 18]. Knots also play a role in biological macromolecules. They were first discovered in DNA in 1976 [19] and some can even be identified in the backbones of proteins [20, 21, 22, 23, 24, 25, 26, 27, 28].

All in all, the study of "physical knots" has become an active interdisciplinary field of research. The intention of this manuscript is to serve as an introductory and by no means exhaustive primer on technical aspects which arise in the computer-aided analysis of simple knots. After providing elementary definitions (chapter 2), we describe the Alexander polynomial and its implementation in a simulation code (chapter 3). In chapter 4 we present a simplification algorithm which reduces the computational effort considerably and allows us to visualize knotted configurations, too. In chapter 5 we discuss how to extend the notion of knottedness to "physical knots" in open chains by applying statistical concepts.

1875-3892
doi:10.1016/j.phpro.2010.09.036

2. Basic technical terms

In mathematics, knots are only well-defined in closed curves embedded in three dimensional space [29, 30]. Most commonly they are classified by counting the minimum number of crossings (the crossing number) in a projection onto a plane. Fig. 1 shows the most basic knots. An unknotted loop (0) is typically named "unknot". The first non-

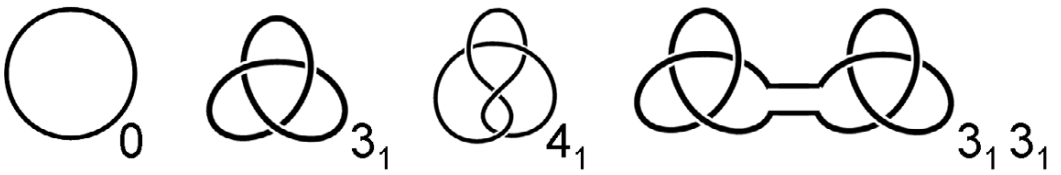

Figure 1: Knots are usually classified according to the minimum number of crossings in a projection onto a plane. An unknotted loop (0) is called "unknot". The simplest non-trivial knot, the "trefoil" knot (3_1), has three crossings, the "figure-eight" knot (4_1) has four. The index distinguishes knots with the same crossing number. (There is one knot with three crossings (3_1), one with four (4_1), two with five (5_1 and 5_2), three with six and so on.) *Right*: a simple composite knot is displayed which consists of two trefoil knots.

trivial prime knot, the so-called "trefoil knot" (3_1), contains three minimum crossings in a projection onto a plane, and the "figure-eight knot" (4_1) four. Fig. 1 also shows a composite knot made up of two trefoils ($3_1 3_1$). The index in the numbering scheme differentiates knots with the same crossing number. (A knot table can, e.g., be found in the book of Rolfson [31] or online [32].) Note that the number of knots increases considerably with the crossing number (see Fig. 2). Therefore, it is not surprising that no algorithm has been discovered up to now which is able to distinguish between all knots.

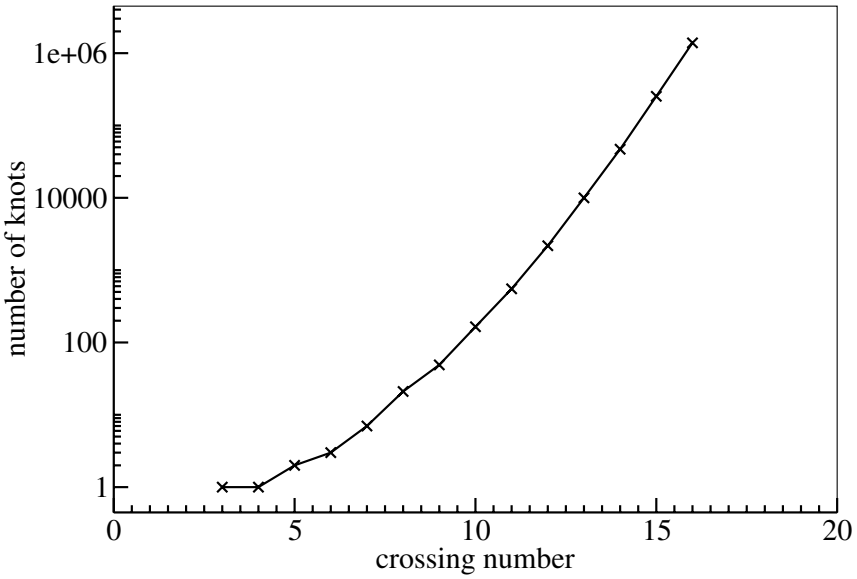

Figure 2: Number of knots as a function of crossing number for knots with up to 16 crossings. Data taken from [33].

A knot invariant is a property of a knot which always assigns the same value to the same knot and can to some extent be used to differentiate between knots. The crossing number, e.g., is already a knot invariant even though it

does a poor job in distinguishing between knots. The first invariant which was somewhat successful in this regard was formulated by Alexander in the 1920s [34]. The so-called Alexander polynomial Δ assigns, as indicated, a polynomial to a knot. Again, Δ is not a complete invariant, i.e, it is not unique for each knot. However, it is sufficient to distinguish between simple knots. For example, unknot, trefoil and figure-eight knot have a unique polynomial among knots with up to crossing number 10. The composite knot $3_1 3_1$, however, coincides with 8_{20} and 10_{140}.

3. Computation of the Alexander Polynomial

3.1. Definition and properties

In this section we define the Alexander polynomial and summarize properties which are relevant for implementing it in a simulation code. The description follows to some degree Ref. [29]. We will, however, focus on algorithmic aspects of the problem. Readers interested in the mathematical background are referred to the relevant introductory literature [29, 30].

In the following we examine a closed loop made up of monomer units which are connected by bonds; open polymer chains will be analyzed in chapter 5. To calculate the Alexander polynomial of such a loop, we consider a projection of the closed three dimensional curve onto a plane (e.g., the x-y plane). First, we number the intersection points and arcs, and choose an orientation of the diagram as exemplified for the trefoil in Fig. 3 *left*.

After selecting an arbitrary starting point (•) and orientation for the diagram, we number the crossings (1,2,3) as we move along our chosen direction. First, we underpass a later section of the curve at crossing 1. At crossing 2 we overpass and at intersection 3 we underpass once again. In this example, arcs are numbered (I,II,III) from underpass to underpass. Note, that this particular assignment is completely arbitrary and any other choice will lead to equivalent results. We also determine the handedness of the crossing (see Fig. 3). Following [29], the arc of the overpassing bond is always assigned to i. The incoming underpassing section is denoted as j, the outcoming underpassing section as k. An intersection is right-handed if the z-component of the cross product between $\vec{i}$ and $(\vec{j,k})$ is positive (assuming the curve was projected onto the x-y plane). Likewise, it is left-handed if the z-component of the cross product is negative.

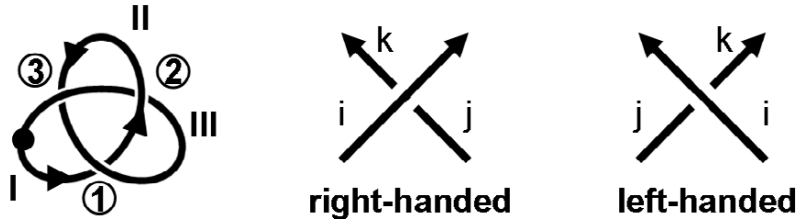

Figure 3: *Left*: exemplary numbering of arcs (I,II,III) and crossings (1,2,3) for a trefoil knot. *Middle* and *right*: definition of handedness of a crossing. The arc of the overpassing bond is always assigned to i. The incoming underpassing section is denoted as j, the outcoming underpassing section as k. An intersection is right-handed if the z-component of the cross product between $\vec{i}$ and $(\vec{j,k})$ is positive (assuming the curve was projected onto the x-y plane). Likewise, it is left-handed if the z-component of the cross product is negative.

Now, we define an $n \times n$ matrix, where n is to the number of crossings. Each row of the matrix corresponds to a particular intersection. The matrix elements are obtained according to the following rules:

- If crossing l is right-handed, assign the following entries to row l:
  A_{li}=1-t, A_{lj}=-1, and A_{lk}=t. (t is a variable.)
- If l is left-handed crossing, assign:
  A_{li}=1-t, A_{lj}=t, and A_{lk}=-1.
- If any two indexes match (e.g., $i = j$), assign
  the sum of the entries above to the appropriate column.

- All remaining entries in row l are set to 0.

The Alexander polynomial Δ is determined by computing the determinant of the $(n-1)\times(n-1)$ matrix obtained from removing the last row and column of this matrix.

EXAMPLE: We calculate Δ for the trefoil depicted in Fig.3 *left*. All crossings are right-handed:
Intersection 1: i=III, j=I, k=II $\Rightarrow$ A_{11}=-1, A_{12}=t.
Intersection 2: i=II, j=III, k=I $\Rightarrow$ A_{21}=t, A_{22}=1-t.
$\Delta = -1(1-t) - t^2 = -t^0 \cdot (1 - t + t^2)$ (compare with table 1).

Two properties are particularly relevant for our implementation:

- Δ may differ by a factor of $\pm t^m$ (m integer) for different choices of projections and labelings even though the underlying loop is the same.
- Δ of a composite knot (see Fig. 1 *right*) is the product of the single knot polynomials.

Furthermore

- Δ is symmetric in t and t^{-1} (up to the normalization factor $\pm t^m$) and
- $\Delta(1) = \pm 1$.

3.2. Implementation

To compute the Alexander polynomial we need to determine and analyze bonds which cross each other in a projection onto an arbitrary plane, e.g., the x-y plane. For each intersection, we identify the overpassing bond by comparing the respective z-components of the bonds at the intersection point. The handedness of the crossing is determined by calculating the cross product between over- and underpassing bond. We also establish the distance along the projected curve between the first monomer and the intersection point for both under- and overpass which allows us to number crossings and arcs in a well-defined manner.

Arcs are defined from underpass to underpass. By comparing the distances from the first monomer to the underpass at each crossing we can assign variables j and $k = j + 1$. Similarly, we can assign the arc of the overpass and obtain variable i. (Note that the arc between the last underpass and the starting point belongs to the first arc, too.) Now, we can set up the matrix as described above and compute its determinant with an LU decomposition, which is ,e.g., implemented in Numerical Recipes [35]. From a practical point of view, it is helpful to store information about intersection points in an array of structs (if C is used) or a similar data structure. Each element contains information about a particular crossing, namely the distance from the first monomer, the handedness of the crossing and regions assigned to i,j and k.

Two more problems need to be addressed. As outlined above, $\Delta(t)$ can only be determined up to a factor of $\pm t^m$ with m being an arbitrary integer. On the other hand, it is difficult to compute numerically the full polynomial using machine algebra as suggested in Refs. [7, 8]. A common solution is to calculate only $|\Delta(-1)|$. Unfortunately, $|\Delta(-1)|$ is a much weaker invariant and even some of the simplest knots like 4_1 and 5_1 cannot be distinguished (see table 1). Instead of calculating additional invariants such as Vassiliev numbers [10, 13] to further discriminate knots, we propose a technical trick [14] which addresses both problems and which is similar to an approach suggested in Ref. [23]. If one calculates $\Delta_p \equiv |\Delta(t) * \Delta(1/t)|$, the prefactors cancel. This product is also a knot invariant which inherits several properties of the original Alexander polynomial. First of all, $|\Delta_p|$ of the unknot is always one. Secondly, $|\Delta_p|$ of a composite knot is the product of $|\Delta_p|$ of the single knots from which it is derived. For our implementation we choose $t = -1.1$ and keep the first five decimal places. We have tested that this choice discriminates simple knots up to 10 crossing as well as the complete Alexander polynomial which suffices for our purposes. The Alexander polynomial for knots of up to five crossings as well as $|\Delta(-1)|$ and $|\Delta_p(t = -1.1)|$ are listed in table 1.

Knot	Alexander polynomial Δ	$\lvert\Delta(-1)\rvert$	$\lvert\Delta_p(t=-1.1)\rvert$
unknot	1	1	1
3_1	$1\text{-}t\text{+}t^2$	3	9.05463
4_1	$\text{-}1\text{+}3t\text{-}t^2$	5	25.09099
5_1	$1\text{-}t\text{+}t^2\text{-}t^3\text{+}t^4$	5	25.45745
5_2	$2\text{-}3t\text{+}2t^2$	7	49.25488
$3_1 3_1$	$(1\text{-}t\text{+}t^2)^2$	9	81.98629

Table 1: Alexander polynomial and knot invariants derived from the Alexander polynomial for knots of up to 5 crossings. The composite knot $3_1 3_1$ is also included.

4. Simplification and visualization

Simplifying the structure offers two advantages. First, it reduces the computational effort of knot detection significantly. As outlined above, the calculation of the Alexander polynomial requires the computation of a matrix determinant, which usually takes of order $O(N^3)$ time. Any reduction in the particle number will therefore enhance considerably the speed of knot detection. In addition, an algorithm which deletes beads which are not essential for preserving the topology of the chain also enables us to point out and visualize knotted regions. In the following we present a simplification algorithm which was first introduced by Koniaris and Muthukumar in 1991 [7, 8]. A variant of the algorithm was later developed independently by Taylor [24]. Therefore, we refer to this scheme as the KMT reduction algorithm.

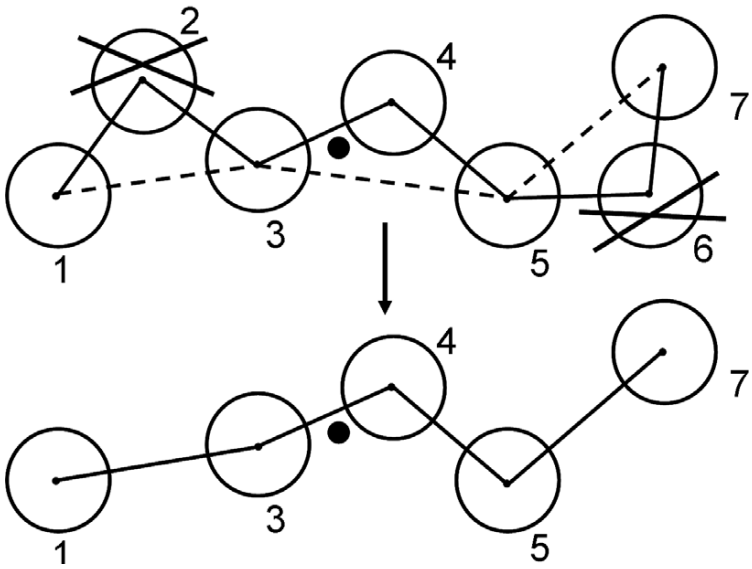

Figure 4: Illustration of the reduction algorithm. We check whether triangles formed by successive monomers are crossed by an impinging bond which belongs to a different section of the chain (black dot). In this example, only monomers 2 and 6 can be removed without crossing bonds.

The simplification algorithm is outlined in Fig. 4. In our implementation, we always keep the first and the last particle. As we move along the chain we check if the chain crosses triangles made up of three successive particles

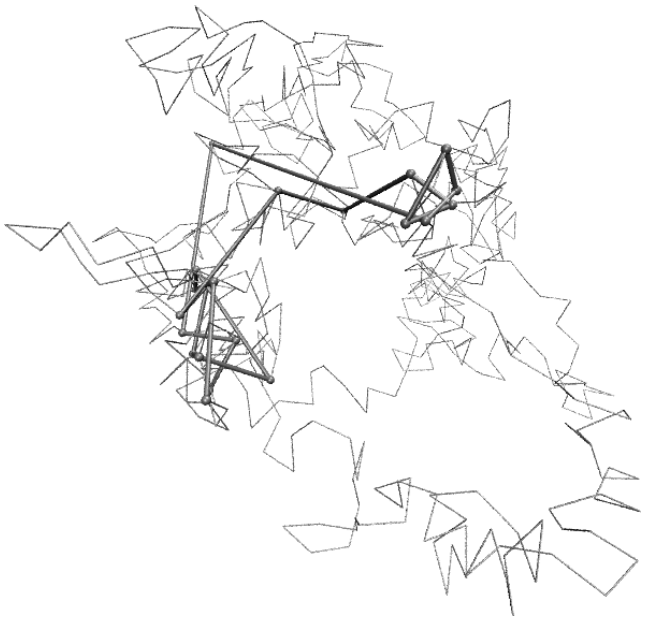

Figure 5: Random loop of size $N = 500$. Thick lines display the structure of the loop after application of the reduction algorithm. The structure contains a small trefoil knot which is located in the upper right corner. The entanglement in the lower right corner is not part of the knot, but was not removed by the algorithm.

(e.g., particles 1,2 and 3 in Fig. 4). If the triangle is not intersected by any part of the remaining chain, we delete the monomer in the middle (#2 and #6). If any bond crosses the triangle (as indicated by the black dot in triangle 3,4,5), we keep the monomer in the middle (#4) because a further reduction could change the relative position of the chain with respect to the intersecting bond. Going back and forth between the two termini we either end up with only these two points or a highly reduced representation of the original chain.

In Fig. 5 we overlay a random loop of size $N = 500$ with its reduced image. Most monomers do not contribute to the topology of the chain and were removed. The structure contains a small trefoil knot which is located in the upper right corner and which can be identified by having a closer look at the structure. Unfortunately, the visualization also has its limitations. In the lower left corner we identify an entanglement which does not contain a knot. In most cases, however, it actually works quite well and provides a helpful tool to point out knotted regions. Nevertheless, it is still important to compute knot invariants to identify knots unambiguously.

Finally, we demonstrate that the simplification algorithm speeds up the computation of the Alexander polynomial considerably. To this extent we apply it to random loops of sizes ranging from $N = 100$ to $N = 500$. Random loops are generated by a generalized "pivot" algorithm: We select two beads at random which subdivide the loop into two segments. Then, one segment is rotated by an arbitrary angle around the connection line between the two points. After repeating this operation 500 times, we obtain a decorrelated loop.

Fig. 6 shows that the number of crossings in a projection onto a plane is reduced by roughly a factor of 17 for all chain lengths considered. This speeds up the computation of the matrix determinant by three orders of magnitude and effectively removes this bottleneck.

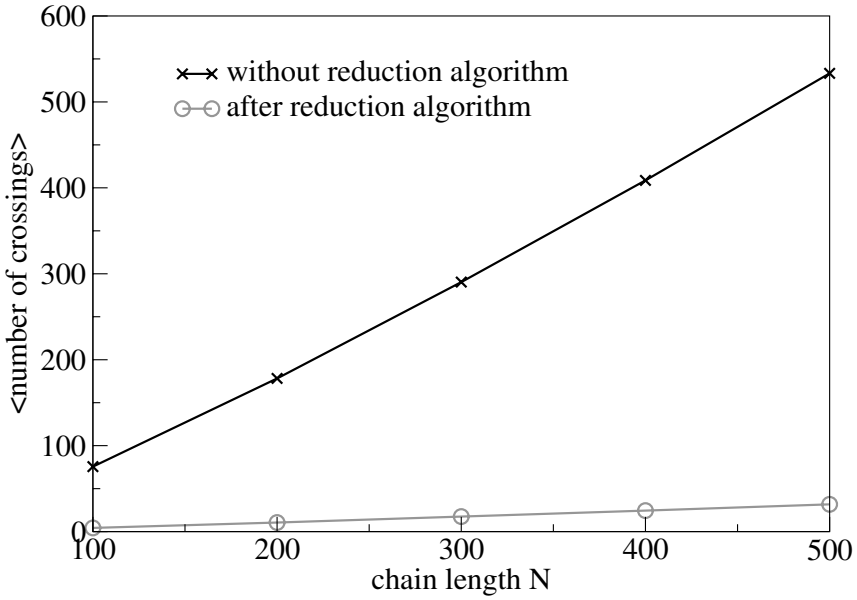

Figure 6: Average number of crossings of a random loop after projection onto the z-plane. Without (x) and after (∘) reduction algorithm. In this case, the number of crossings is roughly reduced by a factor of 17.

5. Closure

As indicated before, a knot is only well-defined in a closed loop. Open chains can always be unknotted (see e.g. Fig.7), but almost all interesting linear macromolecules (polymers, proteins, most DNA strands) are in fact open chains.

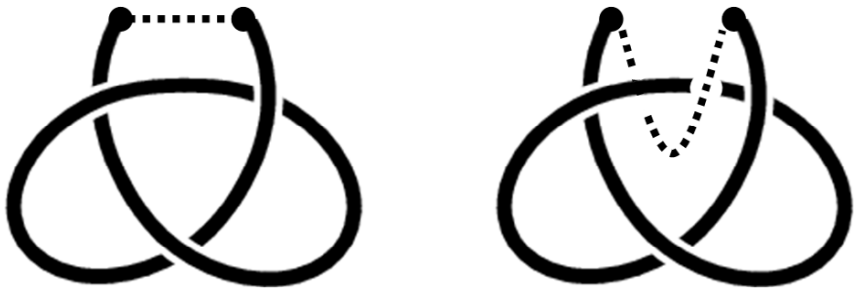

Figure 7: Knots are only well-defined in closed curves. Open strings can always be closed such that they become unknotted (*right*).

That being said, it is possible to expand the concept of knottedness to "physical knots" in open chains in a meaningful manner. For most people - here we exclude mathematicians for a moment - a string is knotted if it does not disentangle after being pulled on both ends. This is equivalent with connecting the end points of an open chain by a big loop which is necessary because the Alexander polynomial is only defined for closed curves. If we apply the same closure to a statistical ensemble of open curves, we can capture the "average" topology and thus information about

the degree of self-entanglements in the ensemble. Note that there are always a few "knots" which only arise due to the closure and which we would consider to be unknotted. However, the influence of these knots on "knot statistics" is minor. If the closure is chosen carefully (closure 3 below), we are even able to preselect single configurations, e.g., particular protein structures, for further analysis [20]. The reliability of the preselection procedure can be improved further by testing not one but several closures on the same structure (closure 2 below) [23, 36, 25].

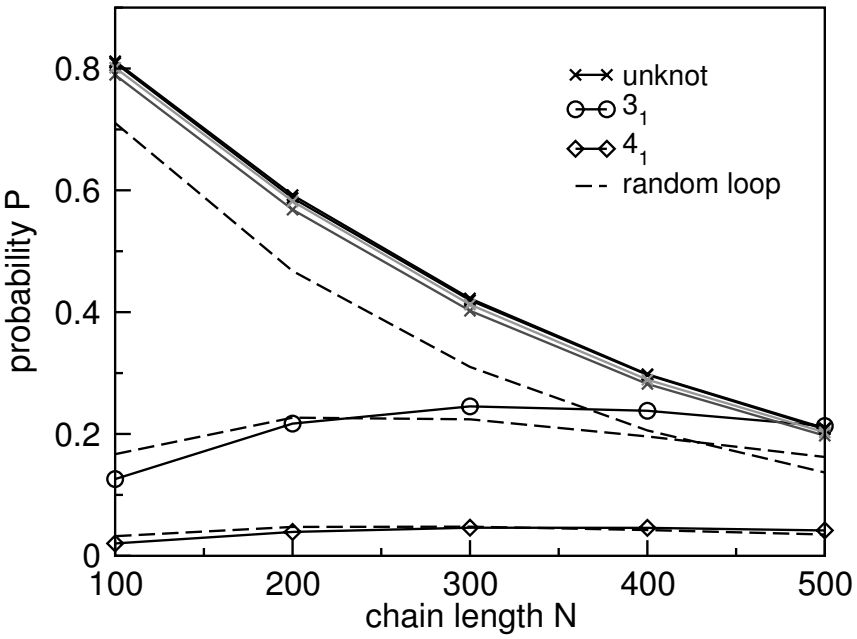

Figure 8: Probability of observing an unknot (×), a trefoil knot (∘) or a figure-eight knot (⋄) in random walks (closed line) and random loops (dashed line) of size 100, 200, 300, 400, and 500. Lines were added to guide the eye. Qualitatively, random loops and random walks yield similar results [36], and the influence of closure is marginal.
For random walks, we applied the KMT reduction algorithm before computing the Alexander polynomial (method 1). For the unknot, alternative closures were evaluated, too. Random closure (method 2) yields results indistinguishable from method 1. For method 3 (see text), a slight decrease of unknotted configurations can be observed. If we connect end points before applying the KMT reduction (method 4), the probability of obtaining unknotted configurations decreases even further.

In the following we show that the influence of the closure is minor indeed and that open and closed loops have a similar "average" topology. To this end we generate an ensemble of open random walks and loops of size $N = 100, 200, 300, 400, 500$ and analyze them with respect to knots. The random walks are generated by successively attaching point-particles on spheres around their predecessors.

Four closures are considered for the open chains:

- Closure 1: We apply the reduction algorithm from section 4 before connecting the end points by a straight line. Note that the application of the reduction algorithm on open chains can potentially change the knot type as pointed out in Ref. [36]. Consequently, closure 1 and 4 yield slightly different results.
- Closure 2: We choose randomly two (connected) points on a large sphere around the chain and connect each end point to one point on the sphere. We repeat the procedure 100 hundred times and select the majority knot type.

- Closure 3: We determine the center of mass of the chain and draw two straight lines outward starting from the center of mass through the end points. At a large distance away from the chain we choose one point on each line. These points are connected with each other and with the end points.
- Closure 4: We connect the end points by a straight line.

Fig. 8 demonstrates that knots are quite common in random walks and loops of medium size [36]. This is not surprising as it is easy to form entanglements on a local scale. For closed random loops of size $N = 500$, more than 85% of all configurations contain a knot. For open random walks the situation is similar. Roughly 80% of all configurations are knotted. Differences between closures are minor.

Acknowledgements

I would like to thank Mehran Kardar, Leonid Mirny, Yacov Kantor, Grisha Kolesov, Daniel Bölinger and Hsiao-Ping Hsu for insightful discussions and fruitful collaborations on the topic of knots and Daniel Bölinger and Thomas Neuhaus for proof-reading. In particular, I would like to thank Kurt Binder for providing a stimulating research environment without which this work would not have been possible. Funding from Deutsche Forschungsgemeinschaft (Vi237/1-1, SFB 625 - project A17) and the Jülich Supercomputing Center is also acknowledged.

[1] W.T. Thompson: Philos. Mag. **34**, 15 (1867)
[2] T.G. Tait: 'On knots I,II and III'. In: *Scientific Papers Vol. 1*, (Cambridge University Press: Cambridge 1898) pp. 273-347
[3] P. Virnau: 'Knots in Macromolecular Systems: Concepts and Challenges'. In: *From Computational Biophysics to Systems Biology, NIC Series Vol. **36***, ed. by U.H.E. Hansmann, J. Meinke, S. Mohanty, O. Zimmermann (John von Neumann Institute for Computing: Jülich 2007) pp. 287-289
[4] H.L. Frisch, E.J. Wassermann: J. Am. Chem. Soc. **83**, 3789 (1961)
[5] M.Delbrück: 'Knotting problems in biology'. In: *Mathematical Problems in the Biological Sciences*, ed. R. E. Bellman, Proc. Symp. Math. 14, (American Mathematical Society, Providence, 1962) pp. 55-63
[6] M.D. Frank-Kamenetskii, A.V. Lukashin, A.V. Vologodskii: Nature **258**, 398 (1975)
[7] K. Koniaris, M. Muthukumar: Phys. Rev. Lett. **66**, 2211 (1991)
[8] K. Koniaris, M. Muthukumar: J. Chem. Phys. **95**, 2873 (1991)
[9] E.J. Janse van Rensburg, D.W. Sumners, E. Wassermann, S.G. Whittington: J. Phys. A: Math. Gen **25**, 6557 (1992)
[10] T. Deguchi, T. Tsurusaki: Phys. Lett. A **174**, 29 (1993)
[11] A.M. Saitta, P.D. Soper, E. Wassermann, M.L. Klein: Nature **399**, 46 (1999)
[12] V. Katritch, W.K. Olson, V. Vologodskii, J. Dubochet, and A. Stasiak: Phys. Rev. E **61**, 5545 (2000)
[13] N.T. Moore, R.C. Lua, A.Y. Grosberg: Proc. Natl. Acad. Sci. U.S.A. **101**, 13431 (2004).
[14] P. Virnau, Y. Kantor, M. Kardar: J. Amer. Chem. Soc. **127**, 15102 (2005)
[15] M. Mansfield: J. Chem. Phys. **127**, 244902 (2007)
[16] D. Reith, P. Virnau: Comput. Phys. Commun. **181**, 800 (2010)
[17] J.-P. Sauvage, C. Dietrich-Buchecker: 'Molecular knots - from early attempts to high-yield template synthesis'. In: *Molecular Catenanes, Rotaxanes and Knots, A Journey Through the World of Molecular Topology*, ed. J-P. Sauvage, C. Dietrich-Buchecker (Wiley-VCH, Weinheim, 107-142, 1999), pp. 107-142
[18] O. Lukin, F. Vögtle: Angew. Chem. Int. Ed. **44**, 1456 (2005)
[19] L.F. Liu, R.E. Depew, J.C. Wang: J.Mol.Biol. **106**, 439 (1976)
[20] P. Virnau, L.A. Mirny, M. Kardar: PLoS Comp. Biol. **2**, 1074 (2006)
[21] D. Bölinger, J. Sulkowska, H.-P. Hsu, L.A. Mirny, M. Kardar, J.N. Onuchic, P. Virnau: PLoS Comp. Biol. **6**, e1000731 (2010)
[22] G. Kolesov, P. Virnau, M. Kardar, L.A. Mirny: Protein knot server: Nucl. Acids Res. **35**, 425 (2007).
[23] M. Mansfield: Macromolecules **27**, 5924 (1994)
[24] W.R. Taylor: Nature **406**, 916 (2000)
[25] R.C. Lua, A.Y. Grosberg: PLoS Comp. Biol. **2**, e45 (2006)
[26] S.E. Jackson, A.L. Mallam: J. Mol. Bio. **346**, 1409 (2005)
[27] W.R. Taylor: Comput. Biol. Chem. **31**, 151 (2007)
[28] R. Potestio, C. Micheletti, H. Orland: PLoS Comp. Biol. **6**, e1000864 (2010)
[29] C. Livingston: *Knot Theory*, (Mathematical Association of America, Washington, 1993)
[30] C.C. Adams: *The knot book: An elementary introduction to the mathematical theory of knots*, (Freeman, New York, 1994).
[31] D. Rolfson: *Knots and Links*, (AMS Chelsea 2003)
[32] http://www.math.toronto.edu/˜drorbn/KAtlas/Knots/ http://www.indiana.edu/˜knotinfo/
[33] J. Hoste, M. Thistlethwaite, J. Weeks: Math. Intell. **20**, 33 (1998)
[34] J.W. Alexander, Trans. Amer. Math. Soc. **30**, 275 (1928)
[35] W.H. Press, S.A. Teukolsky et al: *Numerical Recipes: The Art of Scientific Computing*, 3rd edn. (Cambridge University Press, 2007)
[36] K. Millett, A. Dobay, A. Stasiak: Macromolecules **38**, 601 (2005)