



## OPEN

Subknots in ideal knots, random knots,  
and knotted proteinsEric J. Rawdon<sup>1</sup>, Kenneth C. Millett<sup>2</sup> & Andrzej Stasiak<sup>3,4</sup>

## SUBJECT AREAS:

APPLIED MATHEMATICS  
COMPUTATIONAL SCIENCE

Received

2 November 2014

Accepted

10 February 2015

Published

10 March 2015

Correspondence and  
requests for materials  
should be addressed to  
E.J.R. (ejrawdon@  
stthomas.edu)

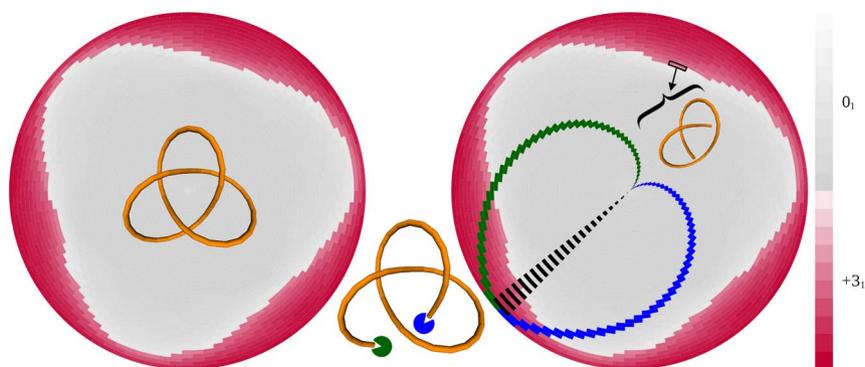
<sup>1</sup>University of St. Thomas, Department of Mathematics, Saint Paul, MN, USA, <sup>2</sup>University of California Santa Barbara, Department of Mathematics, Santa Barbara, CA, USA, <sup>3</sup>University of Lausanne, Center for Integrative Genomics, Faculty of Biology and Medicine, Lausanne, Switzerland, <sup>4</sup>Swiss Institute of Bioinformatics, CH-1015, Lausanne, Switzerland.

We introduce disk matrices which encode the knotting of all subchains in circular knot configurations. The disk matrices allow us to dissect circular knots into their subknots, i.e. knot types formed by subchains of the global knot. The identification of subknots is based on the study of linear chains in which a knot type is associated to the chain by means of a spatially robust closure protocol. We characterize the sets of observed subknot types in global knots taking energy-minimized shapes such as KnotPlot configurations and ideal geometric configurations. We compare the sets of observed subknots to knot types obtained by changing crossings in the classical prime knot diagrams. Building upon this analysis, we study the sets of subknots in random configurations of corresponding knot types. In many of the knot types we analyzed, the sets of subknots from the ideal geometric configurations are found in each of the hundreds of random configurations of the same global knot type. We also compare the sets of subknots observed in open protein knots with the subknots observed in the ideal configurations of the corresponding knot type. This comparison enables us to explain the specific dispositions of subknots in the analyzed protein knots.

Studies of 3-D trajectories of polypeptide chains forming knotted proteins reveal that more complex knots frequently contain simpler knots and slipknots<sup>1–3</sup>. For example, some subchains of a static configuration polypeptide chain forming the  $6_1$  knot can be classified as forming the  $4_1$  knot, while a polypeptide chain forming the  $5_2$  knot has subchains forming  $3_1$  knots<sup>3</sup>. It seems reasonable that as a portion of a knotted chain is shortened, the associated knot type should be progressively simplified until reaching the unknot,  $0_1$ . However, why subchains of  $6_1$  knots should form  $4_1$  knots and subchains of  $5_2$  knots should form  $3_1$  knots is much less evident. Here we study the question: “What are the knot types of the subchains that are contained in a configuration of a complex knot type?” We call the knot types arising from subchains *subknots* of the configuration. Although this question was stimulated by studies of linear knots formed by the polypeptide chains of knotted proteins, we study it here for subknots formed in two special classes of closed chains: the KnotPlot chains [Scharein, R. G. *KnotPlot*. (1998) Available at: <http://www.knotplot.com/>] which visually reflect the structural regularity of the classical prime knot presentations and preserve the knot types’ symmetries<sup>4,5</sup> and the ideal knot configurations<sup>6–13</sup> whose structural properties reflect the spatial nature of knotted magnetic flux lines and of knotted macromolecules<sup>6,14–21</sup>. We compare the sets of subknots to the knot types obtained by changing crossings in minimal knot diagrams for the knot types, the so-called predecessor knots<sup>22,23</sup>. We then compare the subknots seen in these regular configurations to the subknots seen in random configurations. Building upon this study, we consider linear polypeptide chains and discuss what the resulting information tells us about the presence of certain knotted subchains within a knotted polypeptide chain.

### Results and Discussion

**The disk matrix reporting the knot type of every subchain in a closed chain.** Taylor and later King et al. introduced square and triangle-shaped matrices in which the cells report the knot types of the subchains of a linear chain<sup>2,24</sup>. This type of matrix, however, does not adequately reflect the circular periodicity found in a closed chain and, thus, is not well-suited to report the knot types of subchains for closed, circular chains. To overcome this problem, we introduce a disk matrix (see Fig. 1) that reports the knot type of every subchain in a fixed embedding of a circular polygonal chain. It is helpful to think of the disk matrix as being composed of cells delimited by longitude lines radiating from the center of the disk and concentric latitude circles with increasing radius. The matrix cells close to the center of the disk represent very short subchains (starting from one segment), whereas cells bordering the rim of the disk represent long subchains (missing just one segment). The longitudinal position of a cell indicates the position of the midpoint of the associated subchain and the latitude indicates the



**Figure 1** | A guide to disk matrices reporting the knot type of every subchain in a polygonal chain forming a KnotPlot trefoil knot. The left panel shows the disk matrix obtained for the polygonal axial trajectory of a symmetric configuration of a trefoil knot (shown in the center of the matrix). Similar matrices for other knots are presented in Figs. 2–6. The right panel and the drawing between the two panels are intended to explain the principle of the matrix. For the explanation see the main text. The underlying brick wall pattern of the matrix is a consequence of the fact that the longitudinal position of the cell indicates the position of the center vertex of the represented chains. For subchains with even numbers of segments, the centers of these subchains fall at a vertex. For subchains with odd numbers of segments, the centers lie at the middle point of a segment.

length, in number of segments, of the associated subchain. A chain composed of 100 segments, for example, has a matrix with 99 latitude and 100 longitude lines where the Greenwich longitude (i.e. positive  $x$ -axis) corresponds to subchains whose center point is the first vertex of the polygonal chain. The numbering of longitude lines goes in a counter-clockwise direction in our matrix. Colors of the cells in the matrix indicate the dominant knot type of the corresponding subchains, i.e. the knot type most frequently resulting from a uniform closure procedure of the open chain (see Refs. 3, 25–28 and the Materials and Methods section). The intensity of a given color reflects how frequently this dominant knot type occurs among the tested closures<sup>3,28</sup>. Fig. 1 shows the disk matrix reporting the two knot types occurring in subchains of the symmetric trefoil knot configuration shown in the center of this disk matrix. The polygonal trefoil knot,  $3_1$ , consists of the center-lines of 47 cylindrical segments. This trefoil configuration has a three-fold rotational symmetry that one also can see in the symmetry of the disk matrix. Near the center of the matrix, we have entries corresponding to short subchains. These entries are colored gray (see the color scale at the right of Fig. 1) indicating that the dominant knot type is the unknot,  $0_1$ , for the closures of these short subchains. As one moves away from the center and closer to the edge of the disk matrix, the individual cells represent subchains that have sufficient length to form trefoil knots as the dominant knot type upon closure. These cells are colored red to indicate the trefoil knot. Notice that cells close to the border between the zones of the  $3_1$  knot and the unknot have colors of decreased intensity (red and gray, respectively). This border effect indicates that the corresponding subchains show a decreasing preference to form the indicated knot types as the closures also create increasing numbers of other knot types, for example those knot types that dominate the other side of the apparent border. Since the knot configuration, a KnotPlot trefoil, is spread out it takes quite a bit of length to realize the global knot type. Later we analyze random configurations, in which the subknots are more localized (i.e. the coloring starts much closer to the center of the disk matrices) and where there is a more diverse spectrum of subknots. Note that we only see unknot and trefoil subchains in this highly regular trefoil knot due to its relatively simple spatial structure.

For each knotted configuration, there is a shortest length at which the global knot is realized. The minimal length subchain or subchains realizing the host knot is called the knot core<sup>29</sup> and is usually determined by the cell(s) closest to the center of the shortest subchain having the global knot type. In the right panel of Fig. 1, one such cell corresponding to a knot core is outlined in black with the corresponding subchain shown nearby. In this panel, the cells colored blue

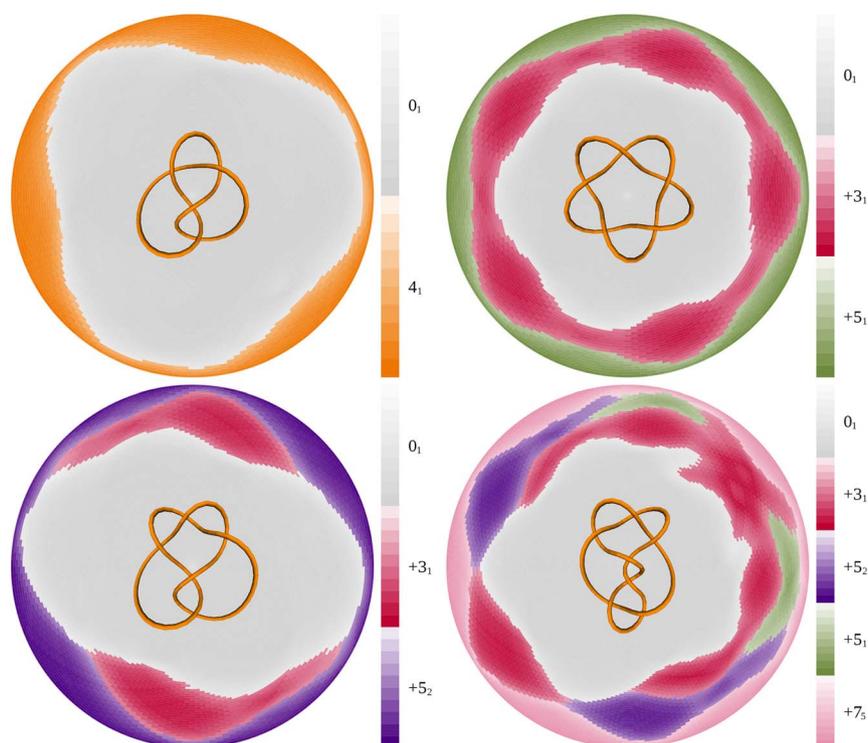
and green represent the subchains resulting from progressively shortening the subchain from each end one segment at a time (represented by the blue and green “pacmen” in the central figure) starting from the same initial scission. The black cells represent the result of simultaneously removing a segment from both ends, thereby shortening the chain by two at each step and giving the dashed pattern. Note that the centers of the chains resulting from progressively removing segments from one end define a spiral path moving from the rim global knot to the center unknot. The direction of the spiral reflects the choice of the end that is being trimmed.

**KnotPlot configuration subknots and predecessor knots.** Fig. 2 shows disk matrices for closed chains forming several other knot types. These knotted chains are configurations created using KnotPlot and are configurations resulting from computations that mimic the action of Coulomb forces on charged elastic fibers forming a given knot type. We chose this group of knot configurations for our initial study because they reflect symmetries of the knot types and the configurations have a projection that looks very similar to the minimal crossing diagrams of the knot types<sup>4,5</sup>. We analyze the polygonal configurations determined by the centerlines of these tubes, taking into account that our tubes are not smooth but composed of many small cylinders.

We continue our analysis with the KnotPlot figure-eight knot,  $4_1$  (Fig. 2A). The figure-eight knot,  $4_1$ , is a twist knot having unknotting number one<sup>5</sup>, as does  $3_1$  (which is both a twist knot and a torus knot). Thus one crossing change can change directly either of them to the unknot<sup>30–32</sup>. This feature is visible in the disk matrix by the direct passage from the global knot type to the gray colored zone as the length of the subchains gets shorter.

Fig. 2B shows the disk matrix for the  $5_1$  knot, another torus knot. Subchains of the  $5_1$  knot are capable of forming the  $3_1$  knot type. Of course, subknots forming the unknot are always observed since any polygon with fewer than six edges (and, thus, subchain with four or fewer edges) is unknotted<sup>33</sup>. The KnotPlot configuration of the  $5_1$  knot has a toroidal five-fold symmetry that can be seen in its disk matrix. The unknotting number of the  $5_1$  knot is two, which also is visible in the disk matrix, since to pass from the green colored  $5_1$  zone to the zone where the subchains only form unknots, one needs to pass through the zone of subchains forming  $3_1$  subknots.

The next example (Fig. 2C) is the  $5_2$  knot, a twist knot (as are  $3_1$  and  $4_1$ ). Twist knots always have unknotting number equal to one. Again, as in the case of disk matrices for the  $3_1$  and the  $4_1$  knots, one can pass directly from the global knot zone to the unknot zone. One also can pass through the  $3_1$  intermediate zone on the way to the zone



**Figure 2** | Disk matrices for KnotPlot configurations of the  $4_1$  (A),  $5_1$  (B),  $5_2$  (C) and  $7_5$  (D) knots. The KnotPlot configurations of the corresponding knots are presented over centers of their disk matrices.

of unknots. The disk matrices we have computed for the KnotPlot configurations (and later ideal knots) with the unknotting number equal to one always showed a direct passage from the zone of the global knot to the zone of the unknot. It is tempting, therefore, to conjecture that this is always the case for knot types with the unknotting number one. The disk matrix of the  $5_2$  knot shows that, as the chain forming the  $5_2$  knot is shortened, it can transition either to a  $3_1$  knot or to an unknot. This resembles the situation in which a minimal diagram of the  $5_2$  knot is subject to single crossing changes<sup>22</sup>. The knots resulting from single crossing changes are either  $3_1$  knots or unknots.

Knot types arising from individual crossing changes performed on a minimal crossing diagram of knot type  $\mathbb{K}$  have been called *predecessors* of  $\mathbb{K}$ <sup>22</sup> since they typically have a smaller minimal crossing number than the knot type  $\mathbb{K}$ . To be more precise, in Ref. 22 the objective was to distinguish predecessors of various generations arising from the classical knot presentations. The first-generation predecessors are the knot types that are obtained by a single crossing change in a minimal crossing diagram of a given knot, whereas the second-generation predecessors are obtained by single crossing changes performed on minimal diagrams of the first-generation predecessors, etc. Diao et al.<sup>23</sup> showed that starting from any minimal diagram of a given alternating knot, one always obtains the same set of first-generation predecessors due to that fact that any such diagram is related to any other by a simple transformation known as a “flype”. For non-alternating knot types, different minimal diagrams can produce different sets of first-generation predecessors. As a consequence, the sets of predecessors for non-alternating knot types depend on the actual knot diagrams chosen and therefore the set of predecessors for non-alternating knot types is not a topological invariant. For this reason, we focus our analysis on alternating knot types that do not have non-alternating predecessors.

The knot  $7_5$  was specifically discussed by Diao et al.<sup>23</sup> and was shown to have  $3_1$ ,  $5_1$ , and  $5_2$  knots as first-generation predecessors. The second-generation predecessors arising from a single crossing change in minimal diagrams of  $5_1$  knots are  $3_1$  knots, those arising

from  $5_2$  knots are  $3_1$  knots and unknots, and those arising from  $3_1$  knots are always unknots. Finally, the third-generation predecessors arising from the  $3_1$  knots that have come from  $5_1$  or  $5_2$  subknots are also unknots. Fig. 2D shows the disk matrix of the KnotPlot configuration of the  $7_5$  knot. We see that the  $3_1$ ,  $5_1$ , and  $5_2$  knots also form first-generation subknots. First-generation subknots can be recognized easily in the disk matrices as having territories that can be accessed directly from the territory of the global knot while advancing radially toward the center of the matrix. We also see that the  $3_1$  knot, in addition to being a first-generation subknot, is a second-generation subknot that arises by truncating subchains forming the  $5_1$  and  $5_2$  knots. Finally, we see that unknots can emerge as second- or third-generation subknots from first- or second-generation subknots, respectively. Interestingly, the disc matrix of the  $7_5$  knot also indicates the predecessor knots which are more likely to appear after randomly changing a crossing in a minimal crossing diagram of the  $7_5$  knot. The  $5_2$  subknots share the longest border with the global knot  $7_5$  and three of the seven crossing changes to the  $7_5$  minimal diagram result in  $5_2$  knots. Meanwhile, the  $3_1$  and  $5_1$  predecessors each appear in two of the seven crossing changes<sup>23</sup>.

Encouraged by the observed degree of agreement between the subknots and the predecessor knots coming from the minimal crossing diagrams, we compared the KnotPlot configuration subknots to the set of predecessors of all knot types with up to 10 crossings for which the set of predecessors is defined (see above). For knot types with up to seven crossings, the set of observed subknots (of all generations) correspond to the set of predecessor knots (of the corresponding generation). However, as the knots increase in complexity, there is an increasing number of cases where one observes subknots that are not present among the set of predecessors as well cases where some of predecessor knots are not present among the subknots (see Table 1). Interestingly, the predecessor knots that are not present among the subknots belong to the predecessors of second and higher generations. We will discuss later how we might find these higher order predecessors within these configurations. The  $8_{10}$  knot (Fig. 3A) is the first example where we see subknots that are not



**Table 1** | Agreement between the sets of predecessor knots and the sets of subknots observed in KnotPlot and ideal configurations with increasing numbers of crossings. For most of the analyzed knots, all observed subknots in the disk matrices of KnotPlot and ideal configurations belong to the set of predecessor knots of the corresponding global knot type. However, as the crossing number increases some of the KnotPlot and ideal configurations have subknots that are not predecessor knots of the global knot type. When one considers majority subknots (i.e. subknots that achieve at least 50% frequency in some subarc using our closure algorithm), then all of these subknots belong to the sets of predecessor knots of the corresponding global knots. If one concentrates on the knot types forming predecessor knots of the first generation then they are visible as subknots in the disk matrices of the KnotPlot and ideal configurations of the corresponding global knots

	number of crossings	3	4	5	6	7	8	9	10
KnotPlot	# alternating knot types with predecessors	1	1	2	3	7	18	35	92
	all subknots $\subseteq$ predecessors	1	1	2	3	7	16	27	58
	some subknots $\not\subseteq$ predecessors	0	0	0	0	0	2	8	34
	all majority subknots $\subseteq$ predecessors	1	1	2	3	7	18	35	92
	first generation predecessors $\subseteq$ subknot set	1	1	2	3	7	18	35	92
	all first generation predecessors $\subseteq$ majority subknot set	1	1	2	3	7	18	32	78
Ideal	some first generation predecessors $\not\subseteq$ majority subknot set	0	0	0	0	0	0	3	14
	all subknots $\subseteq$ predecessors	1	1	2	3	7	18	35	89
	some subknots $\not\subseteq$ predecessors	0	0	0	0	0	0	0	3
	all majority subknots $\subseteq$ predecessors	1	1	2	3	7	18	35	92
	all first generation predecessors $\subseteq$ subknot set	1	1	2	3	7	18	35	92
	all first generation predecessors $\subseteq$ majority subknot set	1	1	2	3	7	18	35	92

predecessors. In addition to the predecessors  $6_3$ ,  $3_1\#-3_1$ ,  $5_1$ ,  $5_2$ ,  $3_1$ ,  $-3_1$  and  $0_1$ , the KnotPlot  $8_{10}$  configuration also contains a  $7_5$  subknot (indicated with an arrow).

**Ideal configuration subknots.** Ideal knot configurations are defined by the axial trajectories of uniform diameter tubes that reach the minimum length necessary to form a given knot type<sup>6–13</sup> and have been shown to have properties that correspond to those found in knotted magnetic flux lines and knotted macromolecules<sup>6,14–21</sup>. Visually, one observes that these configurations are more compact than KnotPlot configurations as a consequence of the minimization of the amount of “rope” used to create the knot.

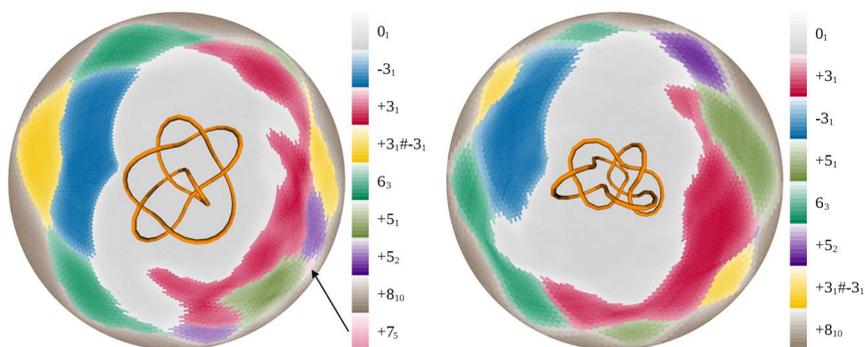
Fig. 3B shows the disk matrix for the ideal  $8_{10}$  knot. All of the predecessor knot types (i.e.  $6_3$ ,  $3_1\#-3_1$ ,  $5_2$ ,  $5_1$ ,  $+3_1$ ,  $-3_1$  and  $0_1$ ) occur while the  $7_5$  knot does not occur.

This result suggests that the reduction of the 3-D trajectory to the necessary minimum required to build a given knot reduces the presence of subknots which are not predecessors. Indeed we observe fewer non-predecessor subknots in ideal configurations than in KnotPlot configurations. However, three 10-crossing knot types ( $10_{69}$ ,  $10_{97}$ , and  $10_{114}$ ) have subknots that are not predecessors. For example, the ideal ( $10_{69}$ ) has a subknot  $7_3$  that is not among the predecessors of that knot. Analyzing this case more closely we noticed that although there are subchains of the ideal  $10_{69}$  configuration that form  $7_3$  knots more frequently than any other knot types upon the uniform closure procedure, the fraction of closures forming  $7_3$  knots is around

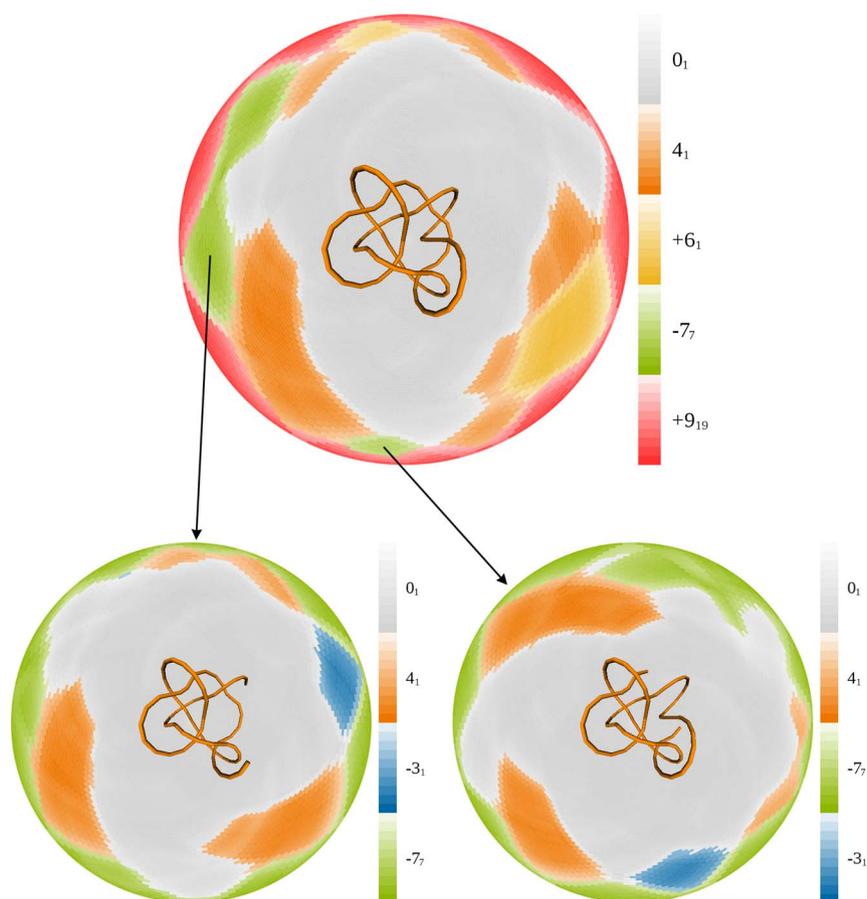
20%. This observation prompted us to consider a more discriminating class of subknots, which we call the *majority subknots*, consisting of knot types that are formed in at least 50% of the closures for some subchain. Interestingly, all of the majority subknots observed in the analyzed ideal and KnotPlot configurations belong to the set of predecessors of the corresponding knot types. We then analyzed whether all predecessor knots are observed among the majority subknots of ideal knots. Some predecessors are not represented amongst the majority subknots but only for predecessors of second and higher generations. All first-generation predecessors are present among the majority subknots of ideal knots. This is not the case, however, for the KnotPlot configurations where some of the first-generation predecessors do not reach the strict criterion of 50% closures (see Table 1).

Among KnotPlot and ideal knot configurations of all prime knots through 10 crossings, only  $3_1$  and  $4_1$  do not contain subknots other than the global knot and the unknot. Furthermore, all KnotPlot and ideal knot configurations contain either a  $3_1$  or  $4_1$  subknot.

**An analysis of second- and higher-generation predecessors and subknots.** In all but one of the ideal configurations of knot types with nine or fewer crossings for which the predecessors are defined, 67 in total, we found that the set of predecessor knots and the set of subknots of the ideal configurations were the same. Fig. 4 shows the disk matrix of the one exceptional case, an ideal  $9_{19}$  knot. This positive  $9_{19}$  knot has a  $-7_7$  knot as one of its first-generation subknots. The predecessors of the  $-7_7$  knot are  $4_1$ ,  $-3_1$ , and  $0_1$ .



**Figure 3** | Comparison of disk matrices for KnotPlot (A) and ideal knot (B) configurations of the  $8_{10}$  knot. Notice that the  $7_5$  knot is visible as one of the subknots in the disk matrix of the KnotPlot configuration (indicated with an arrow), whereas the ideal knot configuration does not contain this subknot. The KnotPlot and ideal knot configurations of the  $8_{10}$  knot are shown over the center of their disk matrices.



**Figure 4 | A procedure that reveals all second-generation predecessors.** Two different subchains that form  $-7_7$  knotted arcs in the ideal configuration of the  $9_{19}$  knot are closed using the underlying idea of closure at infinity whereby two parallel rays are placed at the endpoints of the analyzed subchain. Instead of extending the rays to infinity, the rays are cut as soon they leave the the convex hull of the analyzed subchain and then are closed with an additional segment, yielding a configuration equivalent to the closure at infinity. The corresponding subchains are shown at the center of the corresponding disk matrices. After checking that the closure produced the desired  $-7_7$  knot, the polygonal chains were analyzed to determine their subknots. Note that, in each case, the “extracted”  $-7_7$  knot contains  $-3_1$  subknots even though the  $-3_1$  is not among the subknots of the ideal  $9_{19}$ .

However, the  $-3_1$  subknot is not observed as a second-generation subknot of the ideal  $9_{19}$  configuration.

We are, therefore, led to ask, “Why do the first-generation subknots usually agree with the subknots of ideal configurations but not always those of the second-generation?” This behavior, at least to some extent, comes from differences in the approaches of determining predecessors versus subknots. In particular, predecessors are obtained by a distributive process whereby one crossing is changed in the minimal diagram and then the minimal diagram for the new knot type is analyzed to find the second-generation predecessors. This process is akin to changing one crossing and then changing any other crossing. On the other hand, the analysis of subknots only looks at subchains that are obtained by further trimming subchains that form the given subknot. Thus, the subknot search can be thought of as being a processive process since removing subarcs of increasing length behaves like removing nearby crossings in an ordered fashion as one moves through the configuration. The distributive and processive processes differ in important ways. For example, one does not investigate the subknots that could be revealed if the chain were trimmed at two different portions of the knot. Of course, we cannot open the chain at two (or more) different places using the uniform closure technique<sup>3,28</sup> because there would be four (or more) endpoints of the chain.

To simulate the distributive process, we analyzed the ideal configuration of the  $9_{19}$  knot to see if we could find the second-generation  $-3_1$  knot that emerges from the first-generation  $-7_7$  predecessor.

We took one representative  $-7_7$  knotted subarc from each of the two regions of the  $9_{19}$  that were shown to be  $-7_7$  knots. The regions and the configurations are seen in Fig. 4. We then closed each of the two configurations in one of the closure directions that yields a  $-7_7$  knot and did our subarc analysis on these configurations. We see that both configurations indeed contain  $-3_1$  subknots. We used this procedure to search for eight different second- and higher-generation predecessors that did not appear as subknots in the disk matrices for ideal configurations. In each case the distributive process, such as the one shown in Fig. 4, revealed the predecessors as subknots of lower order subknots.

**Analysis of the subknots found in random configurations of a given knot type.** With the examples above, we have developed an understanding of subknots arising from the classical knot projections, from KnotPlot knots, and from ideal knots. We now ask: “In random configurations, is there a common set of subknots for a given knot type? Furthermore, is the set of subknots related to the set of subknots and predecessors from our previous analysis?” Of course, in the case of random configurations, we expect many different subknots, but could there be a common set?

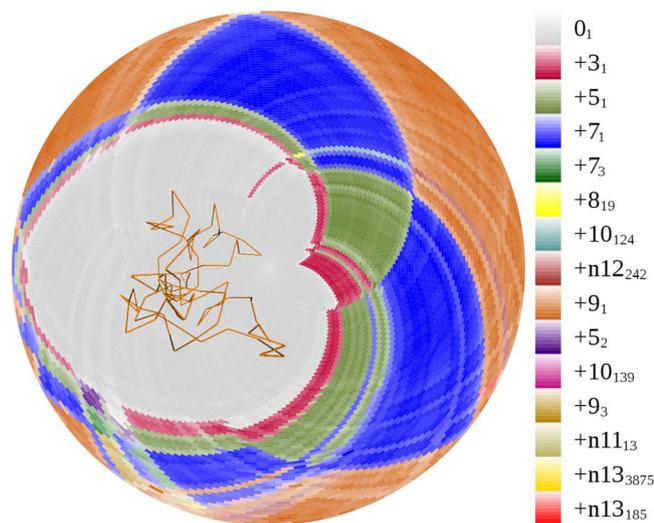
We generated 100,000 random equilateral polygons composed each of 100 segments and analyzed the configurations forming eight or nine crossing knot types that we had analyzed. We chose eight and nine crossing knot types because they have a number of subknots/predecessors and sufficiently large sample sizes.



We start the discussion of random conformations with an analysis of random configurations forming the  $9_1$  knot type. The ideal subknots, KnotPlot subknots, and predecessors are all  $7_1$ ,  $5_1$ ,  $3_1$ , and  $0_1$  and we detected 27 configurations of  $9_1$  knots (right or left-handed). Each of these configurations showed the presence of all of these knot types as subknots although, as expected, a number of additional subknots are also visible. Fig. 5 shows one of these random  $9_1$  knots and its associated disk matrix. We see that the  $7_1$  knot occurs as a first-generation subknot from which  $5_1$  subknots emerge and which, in turn, give rise to  $3_1$  subknots. We also see additional knot types, some of which have a higher minimal crossing number than the global knot. These more complicated subknots frequently arise as subknots in random configurations but appear for only very short intervals of length and are visible only on a small total area of the disk matrix. Furthermore, the more complicated subknots do not appear as subknots of the KnotPlot or ideal configurations and thus are specific to the random configurations instead of being potentially conserved. In the great majority of the configurations of the random non-trivial knots, we see all of the ideal subknots. For example, in each of the 228 configurations of  $8_1$  knots, we always saw the subknots  $6_1$ ,  $4_1$ , and  $0_1$ . And, in each of the 220 random configurations of  $8_2$  knots, we always saw the subknots  $6_2$ ,  $5_1$ ,  $4_1$ ,  $3_1$ , and  $0_1$ .

There are a total of 3334 samples from eight-crossing knot types and 1451 samples from nine-crossing knot types, for a total of 4785 samples. Of these, 4697 ( $\approx 98.16\%$ ) of the samples contained the ideal subknots of all generations. When some of the ideal subknots are not detected in the random configurations, we believe that this often is caused by the fact that the random configurations have many edges that pass very close to other edges. Our analysis will have difficulty detecting all of the transitions in such situations because we only analyze subarcs ending at vertices. Therefore, the removal of a single segment may have the effect of changing several crossings and we can pass directly to higher generation subknots.

While the KnotPlot subknots (90.91%) and predecessors (98.04%) both have good agreement with the subknots of random configurations, the best agreement is with the ideal subknots (98.16%). For the random configurations of some knots types, there are only a few



**Figure 5** | A random  $9_1$  knot and its disk matrix. Notice the presence of the predecessors  $7_1$ ,  $5_1$  and  $3_1$  as first, second, and third-generation subknots, respectively. Notice also the presence of several more complicated subknots, which consume a small amount of the area of the disk matrix. For the knot types with more than 10 crossings, we use the Dowker-Thistlethwaite notation where the leading letter ( $n$  or  $a$ ) tells whether the knot type is non-alternating or alternating, respectively, and the knot types are indexed within the non-alternating and alternating classes.

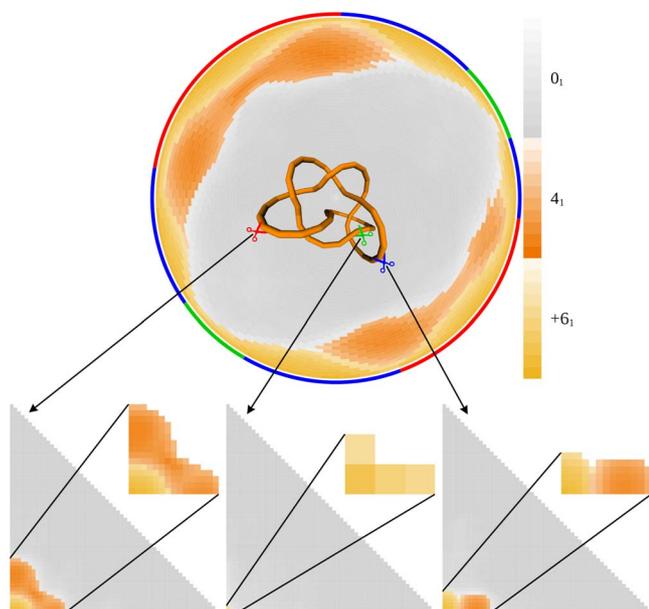
cases where an ideal subknot is not present in a random configuration and we can examine whether refining the image will show the additional knot types. For example, among 228 configurations of  $\pm 8_4$  knots, we found only one configuration that did not show each of the ideal subknots after our standard analysis. When we repeated the analysis after dividing each segment of the same configuration into five equal size segments, we detected the subknot.

There are, however, instances where this finer resolution did not reveal all of the ideal subknots. The most striking case of this phenomena was provided by random configurations of  $8_{16}$  knots, where 15 out of the 64 analyzed configurations did not contain the first-generation  $5_2$  knot. Refinement helped in only one case but not in the remainder of the cases. Therefore, this suggests that at least for some alternating knots, it is possible to have configurations that do not show all of the ideal subknots, and in particular even the ideal first-generation subknots. We note that all 64 analyzed configurations of the  $8_{16}$  knots showed the presence of the remaining first-generation predecessors, i.e.  $6_3$ ,  $4_1$  and  $3_1$ . Thus it is possible that the majority of the predecessors are always present in any random configuration of a given knot type but others may occur less frequently. On the other hand, even if thousands of tested random configurations of some knot type show a common set of subknots, it is possible that there are rare configurations of this knot type that do not contain all of these subknots. While the predecessors and subknots of ideal configurations appear to be good predictors of subknots in random configurations, it is an open question as to whether there even is a common set of subknots for all configurations of a given knot type. If there is a common set of subknots and it does not match the predecessors or ideal configuration subknots, is there some other way to describe them?

#### What can circular chain subknots tell us about protein knots?

Triangle and rectangle-shaped matrices have been used to report the knot type of the subchains of linear chains, e.g. in the analysis of polypeptide chains in searching for knotted and slipknotted proteins<sup>2,3,24</sup>. In the case of more complex proteins structures forming  $6_1$  and  $5_2$  knots, trimming the entire polypeptide from its natural C-terminal-end produces non-trivial subknots  $4_1$  and  $3_1$ , respectively, whereas trimming from the N-terminal end immediately produces unknotted subchains<sup>3</sup>. Does such a difference between directions of trimming also occur in the case of circular knots? If so, what is the position of an opening that would allow trimming in one direction to produce different subknots than trimming in the other direction? Fig. 6A shows the disk matrix for the ideal configuration of the  $6_1$  knot.

Analyzing the disk matrix of the  $6_1$  knot, we see that initial scissions can have three different consequences when the knot is trimmed in the two different directions. There are regions where trimming from either end of an initial scission results in passages from the  $6_1$  global knot to the  $4_1$  subknot before passing to the unknot. These regions are indicated with a red arc near the rim of the matrix in Fig. 6. There are other regions (green) where trimming from either end results in a direct passage from the  $6_1$  global knot to unknots. Finally, there are regions (blue) where trimming of one end results in a passage to the  $4_1$  subknot whereas trimming of the other end results in a passage to the unknot. We also indicate the location of the representative cuts for these three categories of regions in the configuration of this  $6_1$  knot. For a scission in each of these regions, we computed the triangular matrix used in the analysis of knottedness of linear chains such as those formed by the knotted proteins<sup>3</sup>. One can see that one of the three categories of scissions leads to the situation in which trimming of one end creates a  $4_1$  subknot before an unknot is created, whereas trimming of the other end results in a direct passage from a  $6_1$  knot to the unknot (see the right triangular knotting matrix in Fig. 6). Note that the regions where the initial scission followed by trimming of one end results in a  $4_1$  subknot



**Figure 6** | The set of observed subknots in triangular matrices, such as those used to analyze the location of knots in proteins, depends on the position of the segment removed from the analyzed knot. The disk matrix of the ideal configuration of the  $6_1$  knot and three triangular matrices calculated for the linear fragments resulting from cleaving at the indicated spots of this  $6_1$  knot are shown in this figure. The red, green, and blue arcs near the edge of the disk matrix correspond to the regions within the knot from which the initial scissions lead to the formation of linear fragments giving the qualitatively different triangular matrices shown here. Scissions within red arcs give transitions from  $6_1$  to  $4_1$  upon trimming of either end. Scissions within green arcs give direct transitions from  $6_1$  to  $0_1$  upon trimming of either end. Scissions within blue arcs give transitions to the  $4_1$  knot upon trimming of one end and a transition to  $0_1$  upon trimming the other end. The colors of the scissors indicate the color of the arc within which the scission is located.

whereas trimming of the other end results in unknots are placed where the interwound regions end and the apical loops begin. For the ideal  $5_2$  knot, there are also three different regions where initial scissions can result in qualitatively different triangular matrices.

A comparison of the triangular matrices obtained after processive trimming of proteins forming  $6_1$  and  $5_2$  knots (the  $6_1$  case is shown in Fig. 6) provides deep insights into the protein structure. With respect to subknots observed by trimming either end, proteins forming  $6_1$  and  $5_2$  knots produce triangular matrices resembling those observed in ideal knots in which one of the terminal loops is cut so that one end still passes through the other terminal loop. Similar conclusions regarding the location of the polypeptide ends in these proteins have been reached by direct analysis of the knotted protein structures<sup>3</sup>. However, it has not yet been recognized that this particular location of ends is required to achieve the critical property that the trimming of one protein end leads to direct passage to trivial subknots, whereas trimming of other end leads to passages to  $4_1$  or  $3_1$  subknots, respectively, as is the case of proteins forming the  $6_1$  and  $5_2$  knots<sup>3</sup>.

## Methods

The disk matrices for a given polygon are created as follows. For each open arc of the polygon (i.e. consecutive set of edges), we create a number (either 100 or 20) of closed polygons and determine the knot type for each of these closures. The numbers 20 and 100 correspond to the number of closure directions analyzed. To make one closure for an open arc, we create a long line segment (designed to move well outside of the convex hull of the original knot) and place the line segment at both of the arc's endpoints. We then connect the free ends of these added segments to create a closed knot. The 20 directions correspond to the vertices of a dodecahedron. In such a case, the knot type corresponding to each direction is weighted by 0.05 since the dodeca-

hedron is perfectly uniformly distributed. For the 100 directions, we generated a roughly uniform set of points/directions on the unit sphere using Martin's polyhedra<sup>34</sup>. We then computed the Voronoi diagram of that point set on the sphere and weighted each direction based on the percentage of the unit sphere consumed by the Voronoi cell containing the direction. For the ideal, KnotPlot, and random configurations, we used 100 closures. When we refined some of the random polygons for closer analysis, we used 20 closures and then verified the results in the critical regions using 100 closures.

The knot types of the closures are determined using a combination of two techniques. We first use software written by one of the authors (EJR) to encode the crossings of a projection of the polygon. Next the crossing code is simplified (potentially) using Thistlethwaite's unraveller program [Thistlethwaite, M. *unraveller*. (2004) Personal communication]. We then compute the HOMFLYPT polynomial<sup>35</sup> using Ewing and Millett's Impoly program<sup>36</sup>. Note that the HOMFLYPT polynomial is not a perfect knot invariant (i.e. multiple knot types can share a common HOMFLYPT polynomial<sup>35</sup>). We look up the chiral knot types corresponding to the given HOMFLYPT polynomial via a pre-computed table of polynomials. Secondly, we use Thistlethwaite's knotfind algorithm [Hoste, J. & Thistlethwaite, M. *Knotscape*. (1999) Available at: <http://www.math.utk.edu/~morgen/knotscape.html>], which generates a "canonical Dowker code", that we can look up in other tables, to determine a knot type. The knotfind computation is not sensitive to chirality but determines the exact non-chiral knot type. Combining the HOMFLYPT and knotfind information, we can determine the exact chiral knot type for most configurations of knot types with 16 or fewer crossings.

In this study, we analyzed ideal knots, KnotPlot knots, and random knots. The ideal knot configurations were computed using the software Ridgerunner [Cantarella, J., Piatek, M. & Rawdon, E. *Ridgerunner*. (2012) Available at: <http://www.jasoncantarella.com/wordpress/software/ridgerunner/>], a constrained gradient descent algorithm<sup>12</sup> for minimizing the ropelength<sup>9,11,37,38</sup> of a knotted polygon. To reduce the computation time, the ideal configurations from Ref. 12 were reduced (using the splining algorithm splinevect [Cantarella, J. *vecttools*. (2012) Available at: <http://www.jasoncantarella.com/wordpress/software/vecttools/>]) so that their number of vertices was approximately equal to the minimum ropelength. We then used Ridgerunner to tighten the configurations. The KnotPlot configurations are freely available from the software KnotPlot [Scharein, R. G. *KnotPlot*. (1998) Available at: <http://www.knotplot.com>]. The random configurations were created using a variation of the Hedgehog algorithm<sup>39</sup>. We generated 100,000 configurations with 100 edges and the knot types were determined using the techniques described in the previous paragraph.

- Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
- King, N., Yeates, E. & Yeates, T. Identification of rare slipknots in proteins and their implications for stability and folding. *J. Mol. Biol.* **373**, 153–166 (2007).
- Sulkowska, J. I., Rawdon, E. J., Millett, K. C., Onuchic, J. N. & Stasiak, A. Conservation of complex knotting and slipknotting patterns in proteins. *Proc. Natl. Acad. Sci. USA* **109**, E1715–E1723 (2012).
- Rolfsen, D. *Knots and Links* (Publish or Perish, Inc., Texas, 1976).
- Adams, C. C. *The Knot Book* (W. H. Freeman and Company, New York, 1994).
- Katritch, V. *et al.* Geometry and physics of knots. *Nature* **384**, 142–145 (1996).
- Stasiak, A., Katritch, V. & Kauffman, L. H. (eds.). *Ideal knots* (World Sci. Publishing, Singapore, 1998).
- Gonzalez, O. & Maddocks, J. H. Global curvature, thickness, and the ideal shapes of knots. *Proc. Natl. Acad. Sci. USA* **96**, 4769–4773 (1999).
- Litherland, R. A., Simon, J., Durumeric, O. & Rawdon, E. Thickness of knots. *Topology Appl.* **91**, 233–244 (1999).
- Rawdon, E. J. Can computers discover ideal knots? *Experiment. Math.* **12**, 287–302 (2003).
- Cantarella, J., Kusner, R. B. & Sullivan, J. M. On the minimum ropelength of knots and links. *Invent. Math.* **150**, 257–286 (2002).
- Ashton, T., Cantarella, J., Piatek, M. & Rawdon, E. J. Knot tightening by constrained gradient descent. *Experiment. Math.* **20**, 57–90 (2011).
- Cantarella, J., LaPointe, A. & Rawdon, E. J. Shapes of tight composite knots. *J. Phys. A* **45**, 225202 (2012a).
- Stasiak, A., Katritch, V., Bednar, J., Michoud, D. & Dubochet, J. Electrophoretic mobility of DNA knots. *Nature* **384**, 122 (1996).
- Moffatt, H. K. The energy spectrum of knots and links. *Nature* **347**, 367–369 (1990).
- Moffatt, H. K. Helicity and singular structures in fluid dynamics. *Proc. Natl. Acad. Sci. USA* **111**, 3663–3670 (2014).
- Buniy, R. V. & Kephart, T. W. A model of glueballs. *Phys. Lett. B* **576**, 127–134 (2003).
- Maritan, A., Micheletti, C., Trovato, A. & Banavar, J. R. Optimal shapes of compact strings. *Nature* **406**, 287–290 (2000).
- Pieranski, P., Kasas, S., Dietler, G., Dubochet, J. & Stasiak, A. Localization of breakage points in knotted strings. *New J. Phys.* **3**, 10.1–10.13; DOI:10.1088/1367-2630/3/1/310 (2001).
- Weber, C., Carlen, M., Dietler, G., Rawdon, E. J. & Stasiak, A. Sedimentation of macroscopic rigid knots and its relation to gel electrophoretic mobility of DNA knots. *Sci. Rep.* **3**, 1091; DOI:10.1038/srep01091 (2013).



21. Ricca, R. L. Topology bounds energy of knots and links. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **464**, 293–300 (2008).
22. Flammini, A. & Stasiak, A. Natural classification of knots. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **463**, 569–582 (2007).
23. Diao, Y., Ernst, C. & Stasiak, A. A partial ordering of knots and links through diagrammatic unknotting. *J. Knot Theory Ramifications* **18**, 505–522 (2009).
24. Taylor, W. R. Protein folds, knots and tangles. In *Physical and Numerical Models in Knot Theory*, vol. 36 of *Ser. Knots Everything*, 171–202 (World Sci. Publ., Singapore, 2005).
25. Millett, K. C., Dobay, A. & Stasiak, A. Linear random knots and their scaling behaviour. *Macromolecules* **38**, 601–606 (2005).
26. Millett, K. C. & Sheldon, B. M. Tying down open knots: A statistical method of identifying open knots with applications to proteins. In *Physical and Numerical Models in Knot Theory*, vol. 36 of *Ser. Knots Everything*, 203–217 (World Sci. Publishing, Singapore, 2005).
27. Millett, K. C., Rawdon, E. J., Stasiak, A. & Sulkowska, J. I. Identifying knots in proteins. *Biochem. Soc. Trans.* **41**, 533–537; DOI:10.1042/BST20120339 (2013).
28. Rawdon, E. J., Millett, K. C., Sulkowska, J. I. & Stasiak, A. Knot localization in proteins. *Biochem. Soc. Trans.* **41**, 538–541; DOI:10.1042/BST20120329 (2013).
29. Taylor, W. R. A deeply knotted protein and how it might fold. *Nature* **406**, 916–919 (2000).
30. Dazey Darcy, I. & Sumners, D. W. A strand passage metric for topoisomerase action. In *KNOTS '96 (Tokyo)*, 267–278 (World Sci. Publishing, Singapore, 1997).
31. Darcy, I. K. & Sumners, D. W. Rational tangle distances on knots and links. *Math. Proc. Cambridge Philos. Soc.* **128**, 497–510 (2000).
32. Darcy, I. K. Biological distances on DNA knots and links: applications to XER recombination. *J. Knot Theory Ramifications* **10**, 269–294 (2001).
33. Randell, R. An elementary invariant of knots. *J. Knot Theory Ramifications* **3**, 279–286 (1994).
34. Allanson, B. *Martin's polyhedra*. (2012) Available at: <http://members.ozemail.com.au/llan/mpol.html> (Accessed: 4th November 2012).
35. Freyd, P. *et al.* A new polynomial invariant of knots and links. *Bull. Amer. Math. Soc. (N.S.)* **12**, 239–246 (1985).
36. Ewing, B. & Millett, K. C. Computational algorithms and the complexity of link polynomials. In *Progress in Knot Theory and Related Topics*, 51–68 (Hermann, Paris, 1997).
37. Rawdon, E. J. Approximating the thickness of a knot. In *Ideal Knots*, 143–150 (World Sci. Publishing, Singapore, 1998).
38. Rawdon, E. J. Approximating smooth thickness. *J. Knot Theory Ramifications* **9**, 113–145 (2000).
39. Vologodskii, A. V., Anshelevich, V. V., Lukashin, A. V. & Frank-Kamenetskii, M. D. Statistical mechanics of supercoils and the torsional stiffness of the DNA double helix. *Nature* **280**, 294–298 (1979).

## Acknowledgments

This work was supported in part by NSF DMS 1115722 and 1418869 (to E.J.R.), the Swiss National Science Foundation 31003A\_138267 (to A.S.), and the Leverhulme Trust RP2013-K-017 (to A.S.). The authors thank the Institute for Mathematics and its Applications, the Isaac Newton Institute for Mathematical Sciences, and the Banff International Research Centre for hosting their visits.

## Author contributions

K.C.M., E.J.R. and A.S. conceived the project. E.J.R. performed the numerical analysis. K.C.M., E.J.R. and A.S. discussed the results and wrote the manuscript.

## Additional information

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Rawdon, E.J., Millett, K.C. & Stasiak, A. Subknots in ideal knots, random knots, and knotted proteins. *Sci. Rep.* **5**, 8928; DOI:10.1038/srep08928 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder in order to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>